REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL 01 Feb 91 TO 31 Dec 93

4. TITLE AND SUBTITLE
VISUAL MOTION PERCEPTION AND VISUAL INFORMATION PROCESSING

5. FUNDING NUMBERS
AFOSR-91-0178
61102F
2313
AS

2

6. AUTHOR(S)
Dr George Sperling

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dept of Psychology
New York University
6 Washington Place, Room 980
New York, NY 10003

8. PERFORMING ORGANIZATION REPORT NUMBER
AEOSR-TR- 94 0159

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NL
110 DUNCAN AVE SUITE B115
BOLLING AFB DC 20332-0001

Dr John F. Tangney

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFOSR-91-0178

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This final progress report summaries the main recent results; full reports of the results are contained in the papers appended herewith. The summary also reviews some results from previous AFOSR grants where these are necessary to provide the background for the current research. Four areas are summarized:

1. Basic Mechanisms of Visual Motion and Texture Perception
2. Lateral Interations in Texture Stimuli
3. Information Processing
4. Visual Attention and Short-Term Memory.

14. SUBJECT TERMS

15. NUMBER OF PAGES
161

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
(U)	(U)	(U)	(U)

Sperling: Visual Motion Perception and Visual Information Processing

FINAL REPORT AFOSR Grant 91-0178

February 1, 1991 to January 31, 1993 (includes unfunded extention to 31Dec93)

ABSTRACT

This final progress report summaries the main recent results; full reports of the results are contained in the papers appended herewith. The summary also reviews some results from previous AFOSR grants where these are necessary to provide the background for the current research. Four areas are summarized:

1. Basic Mechanisms of Visual Motion and Texture Perception
2. Lateral Interations in Texture Stimuli
3. Information Processing
4. Visual Attention and Short-Term Memory.

1. Basic Mechanisms of Visual Motion and Texture Perception

This project concerned the discovery and description of basic mechanisms of human visual motion and texture perception. Motion and texture are critical i puts to visual perception. Basic mechanisms of motion are of particular interest because they are perhaps the primary substrate for perceptual recovery of 3D depth structures and orientation in space, they are critical for detecting new objects and events in the environment, as well as playing an important role in 2D perception.

Motion and texture are considered together here because the problem of discriminating velocity in a one-dimensional motion stimulus is formally equivalent to the problem of discriminating orientation in a texture stimulus: the t dimension of the motion stimulus becomes the y dimension of the texture stimulus.

First-Order Motion Perception

First-order motion perception. The initial studies, carried out at the inception of AFOSR support, succeeded in describing the basic mechanism of human Fourier motion perception in full mathematical detail. Several critical insights made this possible. The most important was recognizing that the failure of previous theoretical attempts to apply Reichardt (1957) and similar systems models to human vision (e.g. Foster, 1971) was due in large measure to the fact that they had dealt with data obtained with high-contrast visual stimuli. The human motion-processing system behaves in a simple way for stimuli whose contrast is less than about 0.04 to 0.05 (e.g. Nakayama & Silverman, 1985, others). For higher contrasts, early nonlinearities in the visual system make the analysis the motion processing enormously more complex. Additionally, because hundreds of thousands of detectors may contribute to human psychophysical responses, formal models need to explicitly model decision processes. Finally, stimuli needed to be developed that permitted conclusions about basic motion computations independent of the voting/decision rules imposed by higher-order processes.

van Santen & Sperling (1984) was perhaps the first successful application of these basic principles first-order motion perception to humans, principles that are now quite widely accepted. van Santen & Sperling (1985) showed the equivalence of two subsequent models (Adelson & Bergen, Watson & Ahumada) to the van Santen-Sperling version of the Reichardt model and it developed new results.

van Santen, J. P. H. and Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America - A, 1*, 451-473.

The first of two papers by van Santen and Sperling reports that, by elaborating a Reichardt model that had previously been proposed for insect vision, the model gives an excellent account of human psychophysical data for low-contrast stimuli. To apply a Reichardt detector to human vision requires in considering voting rules (e.g., absolute maximum or total power) for detectors because many detectors present possibly conflicting information to the decision stage. There is a full mathematical development of the elaborated theory. Many counter-intuitive predictions were generated by the theory, and three were experimentally tested. (1) A superimposed stationary grating, even of a grating the same spatial frequency as a moving grating, should not adversely affect motion-direction discrimination. (2) Similarly, a stationary flickering grid should have not affect motion discrimination of a moving stimuli with different temporal frequency. When temporal frequencies of the moving and masking stimuli are the same, then anything may happen, even an illusion of motion in the opposite direction. This apparent reversal of direction of the moving grating for certain predicatable phase relations of the masking stimulus was demonstrated experimentally. (3) For certain spatially-sampled displays, the strength of a motion percept is directly proportional to the *product* of the contrast in adjacent regions. All three predictions were verified. These data show that, contrary to "logical intuition," human motion detection does not rely on matching spatial features in successive frames, but rather on matching of temporal sequences in adjacent locations.

van Santen, J. P. H. and Sperling, G. (1985) Elaborated Reichardt detectors. *Journal of the Optical Society of America - A, 2*, 300-321.

This paper extends the predictive power of the elaborated Reichardt model from continuous to two-flash stimuli, and to other displays, such as random dot displays, that had previously been thought to require "feature" models. It points out that the Reichardt model is consistent with a 3D spatiotemporal Fourier analysis of visual displays. However, when complex displays contain several Fourier components of approximately equal perceptual strength, a more complex analysis such as that of the elaborated Reichardt model, is needed to generate predictions. For example, displays in which component Fourier components move in the same direction and at the same temporal frequency exhibit as more convincing movement than displays in which the components move at the same velocity so to preserve 2D rigidity. It was proved that, for elaborated Reichardt detectors, the strength of motion in two flash displays is predicted by separable temporal and spatial components, so that these displays are ideal for studying the pure spatial properties of motion detectors. Finally, it was proved that two alternative computational theories (Adelson & Bergen, 1985 and Watson & Ahumada, 1985) for which no experimental data had yet been generated, were computationally equivalent to the elaborated Reichardt model.

Investigations of Second-Order Motion and Texture

The theoretical analysis and experimental evidence described above establishes an elaborated Reichardt (or equivalent kind of motion computation) as the basic mechanism of motion perception. The work of the current granting period dealt with a newly discovered second mechanism of motion perception, which was called "Second-order" or "Non-Fourier" motion processing to distinguish it from the previously described "First-order" or "Fourier" motion perception. The computational principles that applied to second-order motion perception were found also to apply to the perception of two-dimensional textures.

Chubb, Charles, and George Sperling. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A: Optics and Image Science, 5,* 1986-2006.

This paper sets forth the general principles. It shows how to construct counterexamples to first-order motion computations: visual stimuli which (i) are consistently perceived as obviously moving in a fixed direction, yet for which (ii) Fourier domain energy analysis yields no systematic motion components in any given direction. A general theoretical framework for investigating nonFourier (second-order) motion-perception mechanisms; two central concepts are *drift balanced* and *microbalanced* random stimuli. A random stimulus *S* is *drift balanced* if its expected power in the frequency domain is symmetric with respect to temporal frequency: that is, if the expected power in *S* of every drifting sinusoidal component is equal to the expected power of the sinusoid of the same spatial frequency, drifting at the same rate in the opposite direction. Additionally, *S* is *micro balanced* if the result *WS* of windowing *S* by any space-time separable function *W* is driftbalanced. It is proved that (i) any space/time separable random (or nonrandom) stimulus is microbalanced; (iia) any linear combination of a pairwise independent microbalanced random stimuli is microbalanced, and any linear combination of a pairwise independent driftbalanced random stimuli is driftbalanced if the expectation of each component is zero (a uniform field); (iii) the convolution of independent micro/driftbalanced random stimuli is micro/driftbalanced; (iv) the product of independent microbalanced random stimuli is microbalanced. Examples are provided of classes of driftbalanced random stimuli which display consistent and compelling motion in one direction although they would be completely ambiguous to any first-order motion mechanism. The perception of nonFourier motion stimuli is explained by postulating a linear space-invariant filter followed by a rectifying mechanism that computes (any increasing function of) the absolute value of stimulus contrast followed by Fourier-energy (e.g., Reichardt) motion analysis. All the results and examples from the domain of motion perception are transposable to and illustrated in the space-domain problem of detecting orientation in texture patterns.

Chubb, Charles, and George Sperling. (1989). Second-order motion perception: Space-time separable mechanisms. Proceedings: Workshop on Visual Motion. (March 20-22, 1989, Irvine, California.) Washington, D.C: IEEE Computer Society Press. Pp. 126-138.

This paper shows how various classes of microbalanced displays can be used to derive properties of second-order motion systems. *Microbalanced* stimuli are dynamic displays which do not stimulate mechanisms that apply *standard motion analysis* directly to luminance (e.g., Adelson-Bergen motion-energy analyzers, Watson-Ahumada motion sensors, or elaborated Reichardt detectors.) Because they bypass *first-order* mechanisms, microbalanced stimuli are uniquely useful for studying *second-order* motion perception (motion perception served by mechanisms that require a grossly nonlinear stimulus transformation prior to standard analysis). The paper demonstrates stimuli that are *microbalanced under all pointwise stimulus transformations* and therefore immune to early visual nonlinearities. Such stimuli are used to disable motion information derived from spatial filtering in order to isolate the temporal properties of space/time separable second-order motion mechanisms. They are equally useful to disable the motion information derived from temporal filtering to isolate the spatial properties.

The paper proposes that second-order motion of all of the classes of microbalanced stimuli under consideration can be extracted by a mechanism consisting of the following stages: (1a) band-selective spatial filtering and (1b) biphasic temporal filtering, nonzero in dc, followed by (2) a rectifying nonlinearity and (3) standard motion analysis.

Chubb, Charles, and George Sperling. (1989). Two motion perception mechanisms revealed by distance driven reversal of apparent motion. *Proceedings of the National Academy of Sciences, USA, 86,* 2985-2989.

It is reasonable to ask whether there really are two mechanisms of motion perception or whether one theory can encompasses both. One way to demonstrate the existence of two mechanisms is to stimulate them to simultaneously give opposite outputs in response to the same stimulus. This paper demonstrates two kinds of visual stimuli that exhibit motion in one direction when viewed from near and in the opposite direction from afar. These striking reversals occur because each kind of stimulus is constructed to simultaneously activate two different mechanisms: a short-range mechanism that computes motion from space-time correspondences in stimulus *luminance* and a long-range mechanism whose motion computations are performed, instead, on stimulus contrast that has been full-wave rectified (e.g., the absolute value of contrast). The stimuli were constructed so that half-wave rectification could be excluded. It is concluded that both a Fourier and a nonFourier computation occur. In this and all previously studied cases of 2nd order motion perception, full wave rectification has been shown to be a *sufficient* mechanism; for these stimuli, full wave rectification (versus half-wave rectification) is shown to be *necessary.*

An analogous phenomenon, distance-driven reversal of apparent slant, occurs with texture stimuli. Apparently, in both motion and texture extraction from visual scenes, there are two parallel mechanisms, operating simultaneously, a first-order mechanism that operates directly on the Fourier components of the stimulus, and a second-order mechanism that operates on a spatiotemporally filtered, full-wave rectified transformation of the stimulus.

Chubb, Charles, and George Sperling. (1991). Texture quilts: Basic tools for studying motion-from-texture. *Journal of Mathematical Psychology, 35, 411-442.*

This paper continues the investigation of motion-from-spatial-texture in stimuli that are free from contamination by motion mechanisms sensitive to anything except texture. It offers a formal foundation for some of the results outlined in Chubb & Sperling's (1989) IEEE paper, and reports the results of three demonstration experiments that establish empirical properties of human second-order motion perception. Additionally, some concrete stimulus-construction methods are provided for a special class of random stimuli called *texture quilts.* Although, as is demonstrated experimentally, certain texture quilts display consistent apparent motion, it is proven that their motion content (a) is unavailable to standard motion analysis (such as might be accomplished by an Adelson/Bergen motion-energy analyzer, a Watson/Ahumada motion sensor, or by any elaborated Reichardt detector), and (b) cannot be exposed to standard motion analysis by any purely temporal signal transformation no matter how nonlinear (e.g., temporal differentiation followed by rectification). Applying such a purely temporal transformation to any texture quilt produces a spatiotemporal function P whose motion is unavailable to standard motion analysis: The expected response of every Reichardt detector to P is 0 at every instant in time.

Three quilts were studied experimentally: a quilt that relies on differences in spatial frequency to generate perception of motion, a quilt that relies on sensitivity to differences in orientation, and quilt that relies on the difference between an even texture and a jointly-independent random texture. The simplest mechanism sufficient to sense the motion exhibited by texture quilts consists of three successive stages: (i) a purely spatial linear filter (ii) a rectifier to transform regions of large negative or positive responses into regions of high positive values, and (iii) standard motion analysis. The first quilt demonstrates that the spatial filter is frequency selective. The second quilt demonstrates that there exist orientation selective filters. The third quilt demonstrates that the rectifier cannot embody a perfect squaring (power) function.

Werkhoven, Peter, George Sperling, and Chubb, Charles. (1993). The dimensionality of texture-defined motion: A single channel theory. *Vision Research, 33, 463-485.*

This paper explores texture-defined motion between similarly oriented sinusoidal patches. It exploits two ambiguous motion displays (types I and II) in each of which apparent motion can be perceived in either of two directions. One of these directions is along a homogeneous space-time path in which all successive sinusoidal patches are identical in spatial frequency and contrast. Along the other, oppositely directed, path is composed of heterogeneous patches that vary in spatial frequency and contrast. The striking and counterintuitive result is that for a wide variety of display conditions, perceived motion along the heterogeneous path dominates the homogeneous path. Obviously, when perceived motion along a path composed of alternating high- and low-frequency patches dominates perceived motion along a pure high-frequency path, the strength of texture-defined motion is not governed by a similarity metric.

All the results are explained in terms of an activity transformation. Each patch is assumed to cause a perceptual response (activity). Strength of perceived motion along a path is determined by the product of the activities of adjacent patches along the path. The path with the greatest product dominates.

Whenever a particular combination of patch contrasts and spatial frequencies caused the two motion paths to be balanced in displays of type I, then they were found to be also balanced in type II displays, a condition referred to as *transition invariance*. Under quite reasonable assumptions about the motion mechanism, it was shown that transition invariance implies that activity must be a one-dimensional quantity. Indeed, activity is well-described as the rectified output of a spatial low-pass filter.

Werkhoven, Peter, Charles Chubb, and George Sperling. (1994) Perception of Apparent Motion between Dissimilar Gratings: Spatiotemporal Properties. *Vision Research. (Accepted for publication pending revisions.)*

This paper continues the search for the determinants of the perceptual strength of texture-defined motion (i.e., motion strength of stimuli that have no net directional energy in the Fourier domain). Werkhoven, Sperling, & Chubb (1993) demonstrated that *correspondence* in spatial frequency and contrast between neighboring patches of texture in a spatiotemporal motion path is irrelevant to motion strength, only *activity*—the rectified output of a spatial lowpass filter—mattered. As in Werkhoven et al (1993), the motion stimuli are ambiguous motion displays in which one motion path, consisting of patches of nonsimilar texture, competes with another motion path, having patches only of similar texture. The textural parameters of spatial frequency, contrast, texture orientation (slant), and temporal frequency are systematically explored.

The data show that motion between dissimilar patches of texture (which are orthogonally oriented, have a two octave difference in spatial frequency and differ 50% in contrast) can easily dominate motion between similar patches of texture. The relative motion strengths of two paths is invariant with temporal frequency from 1 to 4 Hz. Analysis of the data shows that the motion computation is largely but not entirely one-dimensional: Extreme orientation differences and very large spatial frequency differences bring into play small but significant contributions of a second dimension (or dimensions).

2. Lateral Interactions in Texture Stimuli: Contrast-Contrast

Chubb, Charles, George Sperling, and Joshua A. Solomon. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences, USA, 86,* 9631-9635.

Various visual illusions that have been demonstrated for first-order stimuli, may be expected to have corresponding second-order illusions. When the illusions are the result of important properties of signal processing, such as boundary enhancement and gain control, the corresponding second-order illusions should be quite informative about the corresponding second-order process.. This paper considers the second-order analog to perhaps the most famous first-order lightness illusion, namely that the apparent lightness of a uniformly illuminated patch depends on the luminance of its surround. Here it is reported that the perceived *contrast* of a test patch P of binary visual noise embedded in a surrounding noise field S depends substantially on the *contrast* of S. When P is surrounded by high-contrast noise, its bright points appear dimmer, and simultaneously, its dark points appear less dark than when P is surrounded by a uniform field, even though local mean luminance is kept constant across all displays. Sinusoidally modulating the contrast P_S of the noise surround S causes the apparent contrast of P to modulate in antiphase to C_S. For P of contrast C_p, nulling procedures show that the induced induced contrast modulation of P reaches 0.45 C_p. This very large, heretofore unnoticed, spatial interaction is unanticipated by all current theories of lightness perception. It suggests a very general principle of perceptual computation: gain control. Gain control may be be a nearly universal process whereby the response of all a detector is normalized relative to the responses of their neighbors in the same and similar classes.

Joshua A. Solomon and George Sperling. (1993). The lateral inhibition of perceived contrast is indifferent to on-center/off-center segregation but specific to orientation. *Vision Research, 33,* 2671-2683.

Chubb, Sperling, and Solomon (1990) showed that the perceived contrast of a test patch of isotropic spatial texture P embedded in a surrounding texture field S, depends substantially on the contrast of the texture surround S. When P is surrounded by a high contrast texture with a similar spatial frequency content, it appears to be have less contrast than when it is surrounded by a uniform field. This paper describes two novel textures: T^+ which is designed to selectively stimulate only the on-center system, and T^-, the off-center system. When the type of C and of S is chosen to be T^+ or T^-, the reduction of C's apparent contrast does not vary with the combination of T^+, T^-. This demonstrates that the reduction of C's apparent contrast is mediated by a mechanism whose neural locus is central to the interaction between on-center and off-center visual systems.

The induced reduction of apparent contrast is :hown to be *orientation specificity*: the reduction of grating C's apparent contrast by a surround grating S, of the same spatial frequency is greatest when C and S have equal orientation. Using dynamically phase-shifting sinusoidal gratings of 3.3, 10 and 20 cpd, the reduction of apparent contrast was measured using different contrast-combinations of C and S.

The results: (1) Both parallel and orthogonal S gratings caused suppression of P's apparent contrast relative to a uniform surround. (2) In all of the viewing conditions, the reduction of apparent contrast induced by the parallel surrounds was at least as great as that induced by the perpendicular surrounds. Often it was much greater (orientation specificity). (3) Orientation specificity increased with greater spatial frequencies and with lower stimulus contrasts. The results suggest a contrast perception mechanism in which both oriented and nonoriented units determine the perceived lightness or darkness of a point in visual space, and every unit is inhibited primarily by similar adjacent units.

3. Information Processing: Frequency Bands, Subsampling, Noise; Space and Object Perception

This cluster of projects determined, in several domains, how to most efficiently package information to an observer. Obviously, issues of external representation of information are inextricably tied to the question of "What internal representation does the observer use?" Such investigations may lead to useful formulations of how to improve both information presentation and observer training. The basic method was to partition the total stimulus information into several spatial frequency bands, and to determine performance individually for the component bands. Additionally, Riedl and Sperling studied cross-band masking and measure how information from component frequency bands combines in a complex, dynamic visual stimulus.

The "Three-stages and two systems" paper in this sequences proposes a theoretical analysis of the basic computations of visual preprocessing. It shows how results from motion and texture discrimination experiments derive from the same mechanisms that serve higher-order object object perception. The eye movement paper in this sequence deals with the internal representation of scenes that derive from a sequence of saccadic eye movements, and with the visual mechanisms that serve the saccadic mode of information acquisition.

Sperling, Wurst & Lu deal with a new method of discriminating early from late attentional filtering of features that occur within at a single location . Their paradigm, which was applied to repetition detection task, is easily be extended to visual search, and this forms the basis of the proposed experiments.

Riedl, Thomas R. and George Sperling. Spatial frequency bands in complex visual stimuli: American Sign Language *Journal of the Optical Society of America A: Optics and Image Science,* 1988, *5,* 606-616.

This project examined dynamic images of individual signs of American Sign Language (ASL) with a resolution of 96 x 64 pixels which were bandpass filtered in adjacent frequency bands. Intelligibility was determined by testing deaf subjects fluent in ASL. (a) It was possible to find four adjacent bands which divided the signal into approximately equally intelligible parts, any one of which yielded adequate identification accuracy (a) By iteratively varying the center frequencies and bandwidths of the spatial bandpass filters, it was possible to divide the original signal into four different component bands of high intelligibility (67-87% for isolated ASL signs). (b) The empirically measured temporal frequency spectrum was approximately the same in all bands. (c) The masking of signals in band *i* by noise in band *j* was found to be proportional to the frequency similarity: $\log |(f_{noise} / f_{signal})\Delta\omega|$ At constant performance, $(RMS)_{signal} / (RMS)_{noise}$ was the same for bands 2, 3, 4 and higher for band 1. (d) The most effective masking noise is slightly lower in spatial frequency than stimulus ($\Delta\omega$=1.4). (e) Intelligibility for the sum of two very weak signals is greater the closer they are in spatial frequency; for strong signals, the reverse is true. The dominant factor for weak signals is square-law additivity of signal power; for strong signals, redundancy within a band is the limiting factor.

Parish, David H. and George Sperling. Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research,* 1991, *31,* 1399-1415.

The 26 upper-case letters of English were used to determine which spatial frequencies are most effective for letter identification, and whether this is because letters are objectively more discriminable in these frequency bands or because observers can utilize the information more efficiently. Six two-octave wide filters produced spatially filtered letters with 2D-mean frequencies ranging from 0.4 to 20 *cycles per letter height.* Subjects attempted to spatially filtered letters in the presence of identically filtered, added Gaussian noise. The percent of correct letter identifications was measured as a function of s/n in each band at each of four viewing distances ranging over 32:1. In this paradigm, *object spatial frequency band* and s/n determine *presence of information* in the stimulus; *viewing distance* determines retinal spatial frequency, and affects only *ability to utilize.* (a) Viewing distance had no effect upon letter discriminability: object spatial frequency, not retinal spatial frequency, determined discriminability. (b) With the assistance of Charles Chubb, an ideal detector was computed for the letter identification task. For these two-octave wide bands, s/n performance of humans and of the ideal detector improved with frequency mainly because linear bandwidth increased as a function of frequency. (c) Human discrimination efficiency (which compares human discrimination to an ideal discriminator) was 0 in the lowest frequency bands, reached a maximum of 0.42 at 1.5 cycles per object, and dropped to about .104

in the highest band. (d) Upper-case letter information is best extracted from spatial frequencies of 1.5 cycles per object height, an with equal high efficiency over at least a 32:1 range of retinal frequencies from .074 to more than 2.3 cycles per degree of visual angle.

Parish, David H., George Sperling, and Michael S. Landy. Intelligent temporal subsampling of American Sign Language using event boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 1990, *16*, 282-294.

This paper investigates the effects of temporal stimulus subsampling and the form of stimulus representation on intelligibility of a complex visual stimulus (American Sign Language). How well can a sequence of ASL frames be represented by a subset of the frames, and how is the subset optimally chosen? Two drastically different representations of frame sequences were investigated: *dynamic* (ordinary video viewing) and *static* (component frames placed side-by-side in a single display). Secondarily, full gray scale images were compared with binary images (cartoons). An *activity-index* was used to select critical frames at event boundaries–moments in the sequence where the difference between successive frames has a local minimum. Identification accuracy (intelligibility) was measured for 32 experienced ASL signers who viewed 84 variously constructed sequences of isolated ASL signs. With dynamic sequences that utilized full gray-scale, activity-index subsampling yielded significantly more-intelligible sequences than simple repetition of every *n*-th frame, achieving relative compression ratios of up to 2:1. For static sequences, activity subsampling with a small, optimal number of frames achieved higher intelligibility than was achieved by choosing every *n*-th frame, for any *n*. Binary images were less intelligible than the gray scale images, and the relative advantage of activity subsampling was smaller.

(1) Event boundaries can be defined computationally. Sequences composed of frames chosen from event boundaries yielded higher intelligibility than sequences composed of equal numbers of frames spaced at regular intervals. (2) Static presentation of subsets of selected frames can yield intelligible ASL "text" of isolated signs and perhaps, eventually, of conversational ASL.

This research opens the general question of how to use printing technology in place of video technology, where the printing technology is enhanced at the point of production by computer graphics techniques. How can an automatically generated sequence of images best be used -- like a comic book -- to represent a dynamic sequence of events. When an artist is required to represent the images for eventual printing, the cost can be prohibitive. When the images can be automatically generated from a video recording, the production costs are minor. The ASL study demonstrates the feasibility of representing a dynamic ASL sign by a simultaneously visible packet of images. Research is needed to determine how these results might be generalized to more complex communications and to practical training problems that involve dynamic actions.

Sperling, George. Three stages and two systems of visual processing. *Spatial Vision*, 1989, *4*, 183-207.

This paper offers a theoretical synthesis of classic work on light adaptation and on visual thresholds for pattern stimuli, work on efficiency of identification in various spatial frequency bands, and work on motion and texture perception, in terms of three stages and two systems of visual processing. The initial question is: How would an internal noise (at various levels of perceptual processing) appear to external observer? This is determined by the internal location of the noise relative to three stages of visual processing: light adaptation, contrast gain control, and a postsensory/decision stage. Dark noise occurs prior to adaptation, determines dark-adapted absolute thresholds, and mimics stationary external noise. Sensory noise occurs after dark adaptation, determines contrast thresholds for sine gratings and similar stimuli, and mimics external noise that increases with mean luminance. Postsensory noise incorporates perceptual, decision, and mnemonic processes. It occurs after contrast-gain control and mimics external noise that increases with stimulus contrast (i.e., *multiplicative* noise). and therefore mimics external multiplicative noise. Dark noise and sensory noise are frequency specific and primarily affect weak signals. Only postsensory noise significantly affects the discriminability of strong signals masked by stimulus noise; postsensory noise has constant power over a wide spatial frequency range in which sensory noise varies enormously. Especially in dealing with modulation transfer functions, there has been considerable confusion over the spectrum of internal sensory noise (which unavoidably depends on spatial frequency) with the gain factor of sensory transmission (which ideally would be independent of spatial frequency).

Two parallel perceptual regimes jointly serve human object recognition and motion perception: a first-order linear (Fourier) regime that computes relations directly from stimulus *luminance*, and a second-order nonlinear (nonFourier) rectifying regime that uses the absolute value (or power) of stimulus *contrast*. When objects or movements are defined by high spatial frequencies (i.e., texture *carrier* frequencies whose wavelengths are small compared to the object size), the responses of high-frequency receptors are *demodulated* by *rectification* to facilitate discrimination at the higher processing levels. Rectification sacrifices the statistical efficiency (noise resistance) of the first-order regime for efficiency of connectivity and computation.

Sperling, George. Comparison of perception in the moving and stationary eye. In E. Kowler (Ed), *Eye Movements and their Role in Visual and Cognitive Processes.* Amsterdam, The Netherlands: Elsevier Biomedical Press, 1990. Pp. 307-351.

This paper reports the construction of an apparatus for producing *simulated* saccades--continuous sequences of images on a stationary retina that are equivalent to the images produced on the retina during saccadic eye movements. Spatial localization was studied for stimuli flashed during real eye movements (using a limbus monitor) and during identical image sequences (simulated saccades) produced on a stationary retina. The comparison between real and simulated saccades gives critical insights into those mechanisms that are particular to saccades. The paper reviews the historically important paradigms (and representative experiments) that purport to deal with special modes of saccadic processing. On the basis of all these data, it proposes a theory to account for saccadic simulation experiments and to deal with such questions about human visual perception as:

Why don't we see the smear produced on the retina during an eye movement?

Why doesn't the world appear to move as a result of the image movements produced by eye movements?

Does the visual system require sudden stimulus onsets (such as those produced by eye movements) to initiate processing episodes?

To serve the perceptual construction of a stable representation of the world, is there a special memory to relate images produced by successive eye movements?

Sperling, G. Wurst, S. A., and Lu, Z-L. (1993). Using repetition detection to define and localize the processes of selective attention. In D. E. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience - A Silver Jubilee* Cambridge, MA: MIT Press. Pp. 265-298.

Can subjects selectively attend to a subset of items in rapid display sequences, when the subset is characterized by an obvious physical feature, but all items occur in the same location. The paradigm is a repetition detection task in which subjects search a very rapidly presented sequence of thirty superimposed frames for an item that is repeated within four frames. Successful detection implies that a match occurs between an incoming item and a recent item retained in short-term visual repetition memory (STVRM). Previous results (Kaufman, 1978, Wurst, 1989) showed that detection of visual repetitions in a rapid stream of items is indifferent to eye of origin and to interposed masking fields, and functions as well for nonsense shapes as for digits. Therefore, STVRM is visual, not verbal or semantic. It is governed by interference from new items; it does not suffer passive decay within the short interstimulus intervals under which it has been tested.

This paper uses a novel elaboration of a repetition detection paradigm. Within the stream, the physical features of the successive items alternate in color, size or spatial frequency. For example, in the size condition, the odd-numbered items in the stream are large and the even-numbered items are small. Subjects attend selectively to *small* (or to *large*) items. Using selective attention instructions with the repetition detection task permits testing the extent to which, at a single location, subjects can filter rapidly-successive items according to their physical characteristics. By presenting all the items at the same location, only attentional selection according to features (and not according to location) is effective. Subjects selectively attended to subsets of characters based on physical differences of orientation, contrast polarity, color, size, spatial bandpass filtering, and polarity-and-size combined.

Results. Efficiency of attentional selection was determined by comparing performance in a stream of characters that alternated a physical feature with performance in two control conditions: One in which the to-be-unattended characters were optically filtered and another in which all characters shared the same physical feature. Selection efficiency in bandpass filtered streams and in the polarity-and-size streams was greater than 50 percent. Attentional selection based on the other physical features was less effective or ineffective.

Corresponding to the benefits of attentional selection in detecting to-be-attended repetitions, there were large costs in the detection of unattended features. Costs were more ubiquitous than benefits.

In addition to studying repetitions of items that shared a physical feature (homogeneous repetitions) heterogeneous repetitions were studied. Costs for detecting heterogeneous repetitions (relative to homogeneous repetitions) were widespread, indicating that physical features are represented in STVRM. The corresponding stimulus benefits of detecting homogeneous repetitions in feature-alternating streams (under equal attention) were small and only occasionally significant.

If the state of attention were represented in STVRM, we would expect a cost in the detection of heterogeneous repetitions with selective attention instructions (because the attentional state would differ for the two elements of the pair). Such costs were observed and, in some instances they occurred even when there was no corresponding benefit for selective attention in homogeneous detections. This was interpreted as a lack of early attentional filtering compensated by a memory tag representing whether or not an item was attended.

Conclusion: The largest attentional effects occur at the level of attentional selection prior to encoding in STVRM (for bandpass and polarity-and-size stimuli) but that, even when early attentional filtering fails, it can still occur in STVRM.

4. Visual Attention and Short-Term Memory

Performance in many visual tasks depends not only on characteristics of the visual system, but also on more cognitive processes involved in processing visual information, such as attention and memory. The experiments seek to dissect the processes involved in short-term attentional control and the corresponding short-term memory systems. The experimental methods mostly involve rapid sequences of displays because our past work has shown that temporal sequences can be used to sample the time course of temporal processing. The work on visual persistence, iconic memory, and related phenomena exemplifies processing in the absence of successive events; i.e., single-event processing.

Background

The attention experiments herein and many prior experiments from the vast literature on visual attention are encompassed in a general theoretical framework. The starting point is the first published demonstration of an attentional operating characteristic (Sperling and Melchner, 1976, 1978a) and the concept of attentional resources developed by Navon and Gopher (1979), Norman and Bobrow (1975), and others.

Sperling, G. A unified theory of attention and signal detection. In R. Parasuraman and D. R. Davies (Eds.), *Varieties of Attention.* New York, N. Y.: Academic Press, 1984. Pp. 103-181. A state of attention is characterized by a particular allocation of processing and mnemonic resources, and this allocation determines the joint performance on two (or more) competing tasks. The Attention Operating Characteristic (AOC) is the range of possible joint performances as resource allocation is varied from one extreme to the other. This paper demonstrates that the AOC is generated by a process that is mathematically equivalent to the process that generates the receiver operating characteristic (ROC) of signal detection theory (i.e., the process partitions observations into either signal or noise response categories).

This article also proposes a formal definition of a task as a triple of two sets (stimuli and responses) and a mapping between them (a utility function). The task definition enabled a distinction between *compound* and *concurrent* tasks. Concurrent tasks were shown to be especially useful in the study of attention, whereas compound tasks involved primarily the study of decision making, and resulted in considerable difficulties when they were applied to attention. The utility function (in the task definition) is essential to understanding human performance. In contemporary, formal theory, "utility" plays the same role as did "purpose" in earlier, informal accounts of behavior.

Sperling, G., and B. A. Dosher. (1986). Strategy and optimization in human information processing. In K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of Perception and Performance. Vol. 1.* New York, NY: Wiley, 1986. Pp. 2-1 to 2-65.

This highly condensed, encyclopedic treatment of a large literature on attention and performance is equivalent to over 200 ordinary book pages plus more than 100 figure panels. Concepts such as formal task definitions, compound and concurrent tasks, attentional resources, attentional operating characteristics, and more generally, strategies to optimize performance, are applied to the interpretation of data from many classical paradigms. This yields a deeper understanding and, in many instances, vastly different conclusions.

Attentional Trajectories.

The Wilson Cloud Chamber and Glaser Bubble Chamber, which are designed to make visible the trajectories of individual atomic and subatomic particles, work by populating the volume within which a particle will move with steam or superheated liquid. When a target particle moves thru the chamber, a few of the molecules it strikes form the nucleus of condensing droplettes or evolving bubbles, and the visible track of these droplettes or bubbles defines the trajectory.

Sperling and Reeves (1980) introduced an analogous procedure in the realm of measurements of human attention. A rapid stream of superimposed visual items was presented at rates of up to 13 per second in a single spatial location. Subjects attended a second location. At a critical moment during the sequence, subjects were cued to execute a shift of attention to the stream location, and to report the earliest four of the items. The historgram (distribution) of the actually reported items (a small fraction of the presented items) defined the rapid growth and subsequent decline of attention at the stream location. This paradigm made it possible to measure reaction times of shifts of visual attention. Indeed, the paradigm allows the measurement not only of the mean reaction time of an attentional shift but of the entire density function of attentional reaction times (ARTs). Mean ARTs were shown to be quite similar to motor reaction times (MRTs) and to covary with MRTs in response to factors such as task difficulty and target predictability.

Reeves, A., and G. Sperling. (1986) Attention gating in short-term visual memory. *Psychological Review, 93,* 180-206.

This paper offers a computational model of a shift of visual attention, greatly enlarging on the procedures of Sperling & Reeves (1980). An attention shift takes attention from its initial location *a* to a second location *b*. While attention is focussed at *a*, stimulus information from *a* is admitted to further processing, and stimulus information from *b* is excluded. After the shift, the roles of *a* and *b* are reversed. The process of shifting attention to *b* is conceptualized as the opening of an attentional gate at *b*. In Reeves and Sperling's (1980) attentional task, location *b* contains a rapid stream of characters, so the attention gate remains open at *b* only for a for a brief period to avoid flooding memory with irrelevant items.

The theory assumes that the fraction of stimulus information passed on to higher mental processes from a location in space and a moment in time is proportional to the attentional allocation at that location. The theory contains only three parameters: First, there is a latency between the signal to shift attention and the start of the attention shift. Second, the time course of gate opening is described by a second-order gamma function with a time constant, typically, of several hundred msec. Third, there is the amplitude of internal noise that determines the signal-to-noise ratio of the internally represented information.

The data set is quite complex, and the theory makes accurate predictions of literally hundreds of data points with these few parameters.

Sperling, George. The magical number seven: Information processing then and now. In William Hirst (Ed.), *The making of cognitive science: Essays in honor of George A. Miller.* Cambridge, UK: Cambridge University Press, 1988.

This article analyzes why the magical number 7 +-2 had such a major impact on cognitive science --it is the most cited experimental/theoretical article in Psychology. The article 7+-2 offers a theoretical account of absolute judgment (sensory categorization) experiments and of short-term memory experiments. Both kinds of experiments have a limit of 7 (bits, and items, respectively). There are no self-citations in the references. All of the evidence Miller used was publically available. Miller, like Sherlock Holmes, was the one who was able to formulate a theory to encompass these data, and it was perhaps the first plausible quantitative theory to deal with the microprocess of cognition.

The second part of the analysis deals with the current status of Miller's proposals. Miller's seven-item limit turns out to depend on factors such as acoustic confusability, implying that the item limit is based on a sensory-based acoustic memory rather than an abstract memory. The review then points out that a single memory system--a stack of seven items--can encompass both the bit and the item limits Miller had proposed. In a sensory categorization experiment, the seven items in working memory are items with-respect-to-which new items are judged. In a short-term recall experiment, they are the to-be-recalled items. Such a stack memory is easily embodied in a neural network. Thus, a simple neural network memory model can encompass the two main tenets of Miller's magical number seven.

Weichselgartner, E., and George Sperling. (1987) Dynamics of automatic and controlled visual attention. *Science, 238,* 778-780.

Uses the Sperling & Reeves (1980) paradigm to isolate and measure the partially concurrent time courses of automatic and controlled attentional shift. The automatic component is extremely rapid, very brief in duration, and relatively effortless. The controlled component has the same time course as the previously measured attention shifts (Sperling & Reeves, 1980; Reeves & Sperling, 1986), is slower, has a longer duration, and is effortful.

Sperling, George, and Weichselgartner, Erich. (199x). Episodic theory of the dynamics of spatial attention. *Psychological Review.* (Under revision.)

This paper re-analyzes previous measurements of visual attention in simple reaction-time, choice reaction-time and complex discrimination experiments in which attention was purported to move continuously across space. All these data plus data from attention gating experiments were shown to be quantitatively predicted by a quantal (episodic) theory of spatial attention that proposes instead: (a) visual attention can be resolved into a sequence of discrete attentional acts (episodes); (b) each attentional episode is defined by its spatial facilitation function $f(x,y)$; (c) the transition at time t_0 between episodes is described by a temporal alerting/gating function $G(t-t_0)$; (d) f and G are space-time separable. In support of the theory, new experiments are reported that use a concurrent motor reaction-time task to assess changes in discriminability with distance. When non-attentional factors are corrected for, the duration of an attention shift is independent of the spatial distance traversed and of the presence or absence of interposed visual obstacles. New experiments that test and confirm the theory are reported.

Gegenfurtner, K. and Sperling, G. (1993). Information transfer in iconic memory experiments. Journal of Experimental Psychology: Human Perception and Performance, 1993, 19, 845-866.

This paper investigates the role of selective and nonselective transfer processes in partial reports of information from briefly exposed letter arrays. In order to report letters, viewers must transfer information from a rapidly decaying persistence trace (iconic memory) to a more durable short term memory. At some time following termination of the display, subjects are cued to report a particular row of letters. Transfer that occurs prior to the cue is nonselective; transfer that occurs after the cue is selective. (a) Performance is unaffected by 10:1 variations in the probabilities of short and long cue delays. This implies that viewers use the same transfer strategies at all cue delays. (b) Information transfer that has occurred at various times t before and after the cue is measured by using a post-stimulus mask at time t to eliminate visual persistence. Nonselective and selective information transfer (before and after the cue) are shown to combine additively. (c) Positions within rows differ substantially in their accuracy of report.

A simple model accounts for partial report (cued) performance at different cue delays both with and without a mask, and for whole report (uncued) performance. (1) The time course of iconic legibility after stimulus termination depends on the retinal location (row). (2) Initial attention is directed to the middle row, subsequently it switches to the cue-designated row. (3) The instantaneous location-specific legibility times the instantaneous state of attention, integrated over time, determines cumulative transfer, subject to the capacity limit of durable storage. A review of earlier computational approaches shows that only this model is capable of giving a self-consistent account of information transfer from iconic memory.

George Sperling: HIP Lab Publications, 1991-93

1991 Landy, Michael S., Barbara A. Dosher, George Sperling, and Mark E. Perkins. Kinetic depth effect and optic flow: 2. Fourier and non-Fourier motion. *Vision Research,* 1991, *31,* 859-876.

1991 Parish, David H. and George Sperling, Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. Vision Research, 1991, *31,* 1399-1415.

1991 Solomon, Joshua A, and George Sperling. Can we see 2nd-order motion and texture in the periphery? *Investigative Ophthalmology and Visual Science,* ARVO Supplement, 1991, *32, No. 4,* 714. (Abstract)

1991 Werkhoven, Peter, Charles Chubb, and George Sperling. Texture-defined motion is ruled by an activity metric--not by similarity. *Investigative Ophthalmology and Visual Science,* ARVO Supplement, 1991, *32, No. 4,* 829. (Abstract)

1991 Sutter, Anne, George Sperling and Charles Chubb, Further measurements of the spatial frequency selectivity of second-order texture meachanisms. *Investigative Ophthalmology and Visual Science,* ARVO Supplement, 1991, *32, No. 4,* 1039. (Abstract)

1991 Chubb, Charles, and George Sperling. Texture quilts: Basic tools for studying motion-from-texture. *Journal of Mathematical Psychology,* 1991, *35,* 411-442.

1991 Chubb, Charles, Joshua A. Solomon, and George Sperling. Contrast contrast determines perceived contrast. *Optical Society of America Annual Meeting Technical Digest, 1991,* Vol. 17. Washington D.C.: Optical Society of America, 1991. P. XX. (Abstract)

1991 Sperling, G. and Wurst, S. A. (1991). Selective attention to an item is stored as a feature of the item. *Bulletin of the Psychonomic Society,* 1991, *29,* 473. (Abstract)

1992 Shih, Shui-I and George Sperling (1992). Cluster analysis as a tool to discover covert strategies. *Proceedings of the Eastern Psychological Association,* 1992, *63,* 41. (Abstract)

1992 Werkhoven, P., Sperling, G., and Chubb, C. (1992). The dimensionality of motion from texture. *Investigative Ophthalmology and Visual Science,* ARVO Supplement, 1992, *33, No. 4,* 1049. (Abstract)

1992 Werkhoven, P., Sperling, G., and Chubb, C. (1992). *Energy computations in motion and texture. Optical Society of America Annual Meeting Technical Digest, 1992,* Vol. 18. Washington D.C.: Optical Society of America, 1992. P. XX. (Abstract)

1993 Sperling, G. Wurst, S. A., and Lu, Z-L. (1993). Using repetition detection to define and localize the processes of selective attention. In D. E. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV: Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience - A Silver Jubilee* Cambridge, MA: MIT Press. Pp. 265-298.

1993 Werkhoven, P., Sperling, G., and Chubb, C. (1993). The dimensionality of texture-defined motion: A single channel theory. Vision Research, 1993, *33,* 463-485.

1993 Solomon, J. A. and Sperling, G. (1993). Fullwave and halfwave rectification in motion perception. *Investigative Ophthalmology and Visual Science,* ARVO Supplement, 1993, *34, No. 4,* 976. (Abstract)

1993 Shih, Shui-I and Sperling, G. (1993). Visual search, visual attention, and feature-based stimulus selection. *Investigative Ophthalmology and Visual Science,* ARVO Supplement, 1993, *34, No. 4,* 1288. (Abstract)

1993 Lu, Zhong-Lin and Sperling, G. (1993). 2nd-order illusions: Mach bands, Craik—O'Brien—Cornsweet. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1993, *34, No. 4*, 1289. (Abstract)

1993 Chubb, C., Darcy, J. and Sperling, G. (1993). Metameric matches in the space of textures comprised of small squares with jointly independent intensities. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1993, *34, No. 4*, 1289. (Abstract)

1993 Sperling, G. (1993). Spatial, Temporal, and Featural Mechanisms of Visual Attention. *Spatial Vision, 7.* 86. (Abstract)

1993 Gegenfurtner, K. and Sperling, G. (1993). Information transfer in iconic memory experiments. Journal of Experimental Psychology: Human Perception and Performance, 1993, 19, 845-866.

1993 Solomon, Joshua A., and Sperling, George. (1993). The lateral inhibition of perceived contrast is indifferent to on-center/off-center segregation but specific to orientation. *Vision Research, 33*, 2671-2683.

1994 Solomon, Joshua A., and Sperling, George. (1994). Full-wave and half-wave rectification in 2nd-order motion perception. *Vision Research, 33.* (In press.)

Papers Under Submission for Publication.

1994 Werkhoven, Peter, Sperling, George, and Chubb, Charles. (1994). Perception of apparent motion between dissimilar gratings: Spatiotemporal properties. *Vision Research, 33.* (Accepted for publication, pending revision)

199x Sutter, Anne, Sperling, George, and Chubb, Charles. (199x). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research, 33.* (Accepted for publication, pending revision)

199x Sperling, George, and Weichselgartner, Erich. (199x). Episodic theory of the dynamics of spatial attention. *Psychological Review, 101.* (Under revision.)

George Sperling: Talks at Symposia and Meetings of Professional Societies

† Indicates an invited address.
* Indicates an abstract of talk was published.

1991 †George Sperling, Helmholtz Club, University of California, Irvine, February 5, 1991. *Dynamics of Visual Attention: Review and a Theory.*

1991 George Sperling, 87th Meeting of the Society of Experimental Psychologists, University of California at Los Angeles, March 16, 1991. *A Theory of Spatial Attention.*

1991 *Solomon, Joshua A, and George Sperling. Talk presented by Joshua A. Solomon. Association for Research in Vision and Ophthalmology, Sarasota, Florida, April 29, 1991. *Can we see 2nd-order motion and texture in the periphery?*

1991 *Werkhoven, Peter, Charles Chubb, and George Sperling. Poster, presented jointly Association for Research in Vision and Ophthalmology, Sarasota, Florida, April 29, 1991. *Texture-defined motion is ruled by an activity metric--not by similarity.*

1991 *Sutter, Anne, George Sperling and Charles Chubb, Poster, presented jointly Association for Research in Vision and Ophthalmology, Sarasota, Florida, May 1, 1990. *Further measurements of the spatial frequency selectivity of second-order texture meachanisms.*

1991 †George Sperling, Neural Networks for Vision and Image Processing. An International Conference Sponsored by BOston University's Want Institute, Center for Adaptive Systems, Tyngsboro, MA 01879, May 11, 1991. *Two Systems of Visual Processing.*

1991 †George Sperling, Neural and Visual Computation Symposium Center for Neural Sciences New York University, NY, May 31, 1991. *The Spatial, Temporal, and Featural Mechanisms of Visual Attention.*

1991 †George Sperling, National Academy of Sciences, National Research Council, Committee on Vision, Conference of Visual Factors in Electronic Image Communications, Woods Hole, MA, July 23, 1991. *Empirical Observations on Image Compression and Comprehension.*

1991 †George Sperling, The International Society for Psychophysics, Washington Duke Inn, Duke University, Durham, North Carolina, New York University, NY October 19, 1991. *The Featural Mechanism of Visual Attention.*

1991 †*Chubb, C., Solomon, J. A. and Sperling, G. Invited paper presented by Charles Chubb. Optical Society of America, San Jose, California November 7, 1991, *Contrast Contrast Determines Perceived Contrast.*

1991 *George Sperling and Stephen Wurst, Paper presented by George Sperling. Psychonomic Society, San Francisco, California November 22, 1991. *Selective Attention to an Item is Stored as a Feature of the Item.*

1992 *Shui-I Shih and George Sperling. Eastern Psychological Association, Boston, Massachusetts, April 4, 1992. *Cluster Analysis as a Tool to Discover Covert Strategy.*

1992 *Werkhoven, P., Sperling, G., and Chubb, C. Association for Research in Vision and Ophthalmology, Sarasota, Florida, May 6, 1992. *The Dimensionality of Motion From Fexture.*

1992 *Werkhoven, W., Sperling, G., and Chubb, C. Optical Society of America, Albuquerque, New Mexico, September 25, 1992. *Energy Computations in Motion and Texture.*

1992 *George Sperling and Hai-Jung Wu, Paper presented by George Sperling. Psychonomic Society, Saint Louis, Missouri, November 15, 1992. *Defining and Teaching Objectively Accurate Confidence Judgments.*

1993 *†George Sperling, Linking Psychophysics, Neorophysiology, and Computational Vision: A Conference to Celebrate Bela Julesz' 65th Birthday. Rutgers University, New Brunswick, NJ. May 1, 1993. *Spatial, Temporal, and Featural Mechanisms of Visual Attention.*

1993 *Solomon, J. A. and Sperling, G. Talk presented by Joshua A. Solomon. Association for Research in Vision and Ophthalmology, Sarasota, Florida, May 4, 1993. *Fullwave and Halfwave Rectification in Motion Perception.*

1993 *Shih, Shui-I and Sperling, G. Talk presented by Shui-I Shih. Association for Research in Vision and Ophthalmology, Sarasota, Florida, May 6, 1993. *Visual Search, Visual Attention, and Feature-Based Stimulus Selection.*

1993 *Lu, Zhong-Lin and Sperling, G. (1993) Talk presented by Zhong-Lin Lu. Association for Research in Vision and Ophthalmology, Sarasota, Florida, May 6, 1993. *2nd-Order Illusions: Mach bands, Craik—O'Brien—Cornsweet.*

1993 *Chubb, C., Darcy, J. and Sperling, G. Talk presented by Charles Chubb. Association for Research in Vision and Ophthalmology, Sarasota, Florida, May 6, 1993. *Metameric Matches in the Space of Textures Comprised of Small Squares with Jointly Independent Intensities.*

1993 *†Sperling, George and Dosher, Barbara A. Talk presented by George Sperling. Linking Psychophysics, Neurophysiology and Computational Vision. A Conference to Celebrate Bela Julesz' 65th Birthday. Rutgers University, New Brunswick, New Jersey, May 1, 1993. *Structure-from-motion: Algorithms, Illusions, Mechanisms.*

1993 †Sperling, George. Geometric Representation of Perceptual Phenomena. A Conference in Honor of Tarow Indow. University of California, Irvine. July 28, 1993. *The Representation of Motion and Texture.*

1993 †Sperling, George. Society for Mathematical Psychology, Twenty-Sixth Annual Meeting, Norman, Oklahoma. Plenary lecture. August 17, 1993. *Second-Order Perception.*

1993 *†Sperling, George. Ciba Foundation Symposium No: 184. Higher-Order Processing in the Visual System. The Ciba Foundation, 41 Portland Place, London, UK. October 21, 1993. *Full-Wave and Half-Wave Mechanisms in Motion and Texture Perception.*

1993 †Sperling, George. International .Workshop on Digital Video for Intelligent Systems. Hosted by Department of Electrical and Computer Engineering, University of California, Irvine, California. December 17, 1993. *An engineering model of human visual processing/Intelligibility of extremely re* duced images.

George Sperling: Invited Lectures at Universities and Institutes

1991 Department of Psychology Colloquium, University of California, Irvine, Irvine, CA, January 10, 1991. *Visual Preprocessing.*

1991 Department of Psychology University of California at San Diego, La Jolla, CA, February 28, 1991. *Mechanisms of Attention.*

1991 University of California, Berkeley Berkeley, California, Joint Cognitive Science Colloquium and Oxyopia Colloquium (Optometry School), March 22, 1991. *Visual Preprocessing.*

1991 University of California, Berkeley Berkeley, California, Department of Psychology/Cognitive Science Colloquium, March 22, 1991. *The Spatial, Temporal, and Featural Mechanisms of Visual Attention.*

1991 Bonny Center for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, CA, April 8, 1991. *Mechanisms of Visual Attention.*

1991 Salk Institute, University of California at San Diego, La Jolla, CA, April 10, 1991. *Visual Preprocessing.*

1991 Department of Psychology, University of Florida at Gainsville, April 26, 1991. *Systems and Stages of Visual Processing.*

1991 Shanghai Institute of Technical Physics, Shangahi, China, June 17, 1991. *How the Human Visual System Computes Visual Motion* [Host: Prof. Kuang, Ding Bo (Director, SITP); Translators: Dr. Zhang, Ming and Chen, Lulin.]

1991 Department of Computer Science, Shanghai Information-Technology Engineers Examination Center, Fudan University, Shangahi, China, June 18, 1991. *Neural Principles of Preprocessing for Human Pattern Recognition.* [Host: Prof. Wu, Lide (Director, SITEEC).]

1991 Department of Electronic Science and Technology, Institute of Applied Electronics, East China Normal University, Shangahi, China, June 20, 1991. *Measuring Attention* and *How the Human Visual System Computes Visual Motion* [Host: Prof. Weng, Moying (Chairman and Director); Translator: Dr. Zhang, Ming.]

1991 Department of Psychology, Beijing University, and Institute of Psychology, Chinese Academy of Sciences, Beijing, China, June 25, 1991. [Host: Prof. Jing, Qicheng (Director, Institute of Psychology)]

Morning: *The Efficiency of Pereception* [Translators: Dr. Zhang, Ken and Prof. Jing, Qicheng.]

Afternoon: *Measuring Attention.* [Translator: Luo, Chun-Rong.]

1991 Computational Vision Laboratory, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China, June 28, 1991. *First- and Second-Order Motion Perception.* [Host: Prof. Wang Shuo-Rong (Director, Institute of Biophysics); Translator: Prof. Wang, Yun-Jiu (Laboratory Director.]

1991 New York University, Cognitive Sciences Colloquium, September 12, 1991. *Is There Attentional Filtering of Items by Feature as Well as by Location?*

1992 Center for Adaptive Systems Boston University, February 25, 1992. *Is There Attentional Selection of Items by Feature as Well as by Location?*

1992 University of Delaware, Department of Psychology Colloquium, March 4, 1992. *Can Visual Attentional Filter Items by Feature?*

1993 University of California, Irvine, Department of Cognitive Sciences, Vision Lunch Series, January 13, 1993. *2nd-Order Motion Perception.*

1993 University of California, Irvine, Bren Fellows Program, Learned Societies Luncheon, UCI University Club, March 9, 1993. *Modeling Mental Microprocesses.*

1993 University of California, Santa Barbara, First Annual Gottsdanker Memorial Lecture (Department of Psychology). May 27, 1993. *A Theory of Spatial Attention.*

1993 Kenneth Craik Club, University of Cambridge, Cambridge, England, October 25, 1993. *Early Visual Processing.*

1993 University of California, Berkeley. December 3, 1993. *A Theory of Spatial Attention.*

George Sperling. Spatial, temporal, and featural mechanisms of visual attention. *Spatial Vision*, 1993, 7, 86.

Spatial, temporal, and featural mechanisms of visual attention

GEORGE SPERLING
Department of Cognitive Sciences, University of California, Irvine CA, USA

Spatial selective attention is determined by an instruction to attend to a location (or set of locations) X, and temporal attention is determined by an instruction to attend during an interval T. Attentional dynamics are studied by instructions to attend first to X1 and then to X2. To measure these forms of attention, x,y,t space is populated with items, and the x,y,t coordinates of all the attentionally-processed items are determined. Results indicate attention can be represented as a sequence of partially-overlapping space-time separable episodes: attention shifts are discrete, not continuous. Each attentional episode i is characterized by a particular spatial distribution of attention f(i; x) and a temporal period g(i; t) during which f(i; x) is effective. The theory applies to many attentional tasks: go/no-go reaction times, choice reaction times, accuracy in cued search, attentional gating paradigms, and partial report. Attention to features (versus attention to a location in space) is studied by presenting a rapid sequence of items containing different features at a single location and requiring attention only to items that contain a particular feature. Featural attention occurs both early (items with unattended features are selectively excluded from memory) and late (attention selectively affects retrieval).

Shui-I Shih and George Sperling, Visual Search, Visual Attention, and Feature-Based Stimulus Selection. Investigtive Opthalmology and Visual Science, 1993, Pg.1288, No.2885.

VISUAL SEARCH, VISUAL ATTENTION, AND FEATURE-BASED STIMULUS SELECTION. *Shui-I Shih*[1] and *George Sperling*[2].

[1]Saint Anselm College, Manchester, NH and [2]University of California, Irvine, CA.

In rapid visual search, does selective attention to the value of a particular physical feature permit early selection of items with that feature value (e.g., red) while items with a different value (e.g., green) are rejected? Or, does a physical feature only direct attention to a location, with subsequent selection being based on locational selection or on complex decision processes? To disentangle the effects of feature selection and location selection, we use a search paradigm that combines *attentional cuing* and *RSVP (rapid serial visual presentation)*. Two dimensions, *size* (small/large) and *color* (red/green), were studied.

Selective attention to different features was jointly manipulated by instructions, presentation probabilities, and payoffs. On each trial, a visual cue indicated the to-be-attended color or size. The subject then searched a rapid sequence of character arrays (6 letters arranged in a circle around fixation) for a single unknown *digit* among the letter distractors. The feature-cue indicated the probability of that feature value in the target digit (e.g., 50%, 80%, 100% red). The noncued dimension (e.g., size) was neutral; the target's spatial location and identity were chosen randomly and independently. The task was to identify and localize the digit and to identify its feature value in the cued dimension.

In Expt 1, *all items in an array had the same feature value, and successive arrays alternated* in feature value (e.g, red/green). Subjects were not more successful in detecting attended-feature targets than unattended targets. In Expt 2, successive arrays also alternated features but the target and exactly one item in every other array had a unique feature value. Now, subjects benefitted from reliable attentional cues. Thus, attentional cues to a physical feature were useful only when they served to direct attention to a spatial location, not otherwise.

Conclusion: In RSVP visual search, *spatial location* is the means by which feature-based stimulus selection is accomplished.

Supported by ONR Perceptual Sciences Program, Grant N00014-88-K-0569 and by AFOSR Life Sciences, Visual Information Processing Program, Grant 91-0178.

J.A. Solomon and G. Sperling. Fullwave and Halfwave Rectification In Motion Perception. Investigtive Opthalmology and Visual Science, 1993, Pg.976, No.1347.

FULLWAVE AND HALFWAVE RECTIFICATION IN MOTION PERCEPTION.

J.A. Solomon[1] and G. Sperling[2]
[1]Syracuse University, NY; [2]University of California, Irvine.

Microbalanced stimuli are dynamic displays that do not stimulate 1st-order motion mechanisms--mechanisms that apply standard (Fourier energy or autocorrelational) motion analysis *directly* to the visual signal. Purpose. To characterize the perceptual transformations of the visual signal, prior to standard motion analysis, that expose the motion of microbalanced stimuli. Methods. Two kinds microbalanced stimuli are tested: (1) *halfwave stimuli*, for which motion information is exposed by halfwave rectification but lost following fullwave rectification and (2) *fullwave stimuli*, for which motion information is exposed by fullwave rectification but lost following halfwave rectification. Additionally, an ordinary squarewave luminance grating was used to stimulate 1st-order mechansims (the *Fourier stimulus*). Results. Given sufficient contrast, both fullwave and halfwave stimuli convey motion. All observers perceive fullwave motion; only 1/3 can perceive halfwave motion. Remarkably, fullwave stimuli are perceived with slightly greater quantum efficiency than Fourier stimuli, and much more efficiently than halfwave stimuli. Tests of motion transparency reveal that when either fullwave and Fourier or halfwave and Fourier gratings are briefly presented simultaneously, there is a wide range of relative contrasts over which the directions of both gratings can be accurately perceived. Conversely, when halfwave and fullwave gratings are added, both motions are perceived, but subjects cannot tell which is which. Conclusions. Motion transparency between Fourier and microbalanced stimuli implies two parallel motion systems. Subjects' failures to discriminate halfwave from fullwave motion in the transparency task suggest that the halfwave system, for those who posses it, is not labeled differently from the fullwave system.

Supported by AFOSR Visual Information Processing Program, Grant 91-0178.

Zhong-Lin, Lu and George Sperling. 2nd-Order Illusions: Mach Bands, Craik-O'Brien-Cornsweet. Investigtive Opthalmology and Visual Science, 1993, Pg1289, No.2889.

2nd-ORDER ILLUSIONS: MACH BANDS, CRAIK–O'BRIEN–CORNSWEET

Zhong-Lin Lu and *George Sperling*, University of California, Irvine, CA 92717.

Previously, Chubb, Sperling & Solomon[1] demonstrated reduction of the perceived contrast of a textured test patch when surrounded by a textured area of a similar spatial frequency delivered to the same eye. Here, we demonstrate manifestations of lateral textural interactions in analogs of luminance illusions in *fullwave random textures*. (A fullwave random texture has constant mean luminance but becomes equivalent to a luminance pattern upon fullwave rectification.) Mach bands are demonstrated in two fullwave textures: First, a random texture in which the contrast of each pixel is chosen randomly and independently to be either +c or -c; second, a texture constructed out of "Mexican hats" (center-surround) micropatterns that are randomly center-light (+c) or center-dark (-c). The Mach band illusion is produced by making c(x) a ramp function of x that varies from c = 0.20 to 0.80 _/‾. An induced band of low contrast is perceived at the bottom of the ramp, and a band of high contrast near the top of the ramp —/‾‾ . These subjective impressions were quantified by using an interleaved staircase procedure to compare the contrast of a vertical slice of the Mach band pattern to an adjacent texture bar that varied in contrast from trial-to-trial. In control sessions, luminance Mach bands were measured similarly The magnitude of the perceptual Mach bands was similar for the luminance stimulus and for the two fullwave textures. Additionally, a fullwave texture stimulus and a luminance stimulus exhibit Craik-O'Brien-Cornsweet illusions of similar magnitudes (the stimulus —╮———╭— looks like —╮____╭—). However, none of these interactions could be demonstrated for halfwave stimuli, i.e., stimuli that become luminance stimuli after halfwave rectification but are neutral to Fourier and to fullwave analyses. Together, these results indicate that the perceptual processes governing 2nd-order spatial interactions, like those governing 2nd-order motion perception, reflect primarily fullwave (versus halfwave) rectification. As in 2nd-order motion perception, 2nd-order spatial processing (after fullwave rectification of the stimulus) is remarkably similar to first-order luminance processing.

[1] Chubb, C., Sperling, G., & Solomon, J. A. (1989). *Proc. Natl Acad Sci, USA 86*. 9631-9635.

Supported by AFOSR Life Sciences, Visual Information Processing Program, Grant 91-0178

The Dimensionality of Motion-from-Texture

Peter Werkhoven,[1] *George Sperling,*[1] and *Charles Chubb.*[2]

[1]Human Information Processing Laboratory, New York University, NY NY 10003, and
[2]Psychology Department, Rutgers University, New Brunswick, NJ 08903.

Texture-defined motion is 2nd-order apparent motion produced by consecutive patches of texture that are constructed to have no useful Fourier motion components.[3] A general model of the perception of texture-defined motion proposes multiple independent nonlinear transformations of the optical input (channels), each channel being followed by standard motion analysis. Here we present an experimental paradigm and a theoretical analysis to determine the dimensionality of (i.e., the number of channels used in) the motion-from-texture computation and their sensitivity to spatial frequency, orientation and contrast.

Each display contains two competing apparent motion paths. Each path consists of alternating patches of different types of texture. The mean luminance of all textures is equal to the background luminance, and their phases are randomized so that motion detection cannot be based on a direct correlation of the luminance patterns (1st-order motion). We demonstrate heterogeneous motion paths in which consecutive textures differ by two octaves in *spatial frequency*, by a factor of two in *contrast*, and have perpendicular *orientations*, and nevertheless these heterogeneous paths may have the same or greater strength of apparent motion than homogeneous motion paths that consist entirely of either type of texture alone. Such striking counterintuitive results obtain for broad ranges of spatial and temporal frequencies. These and similar results, together with our previous results (ARVO 1991), define a unique single-channel computation for the detection of texture-defined motion. That is, all these sinusoidal texture-defined motion stimuli are preprocessed by a *single* nonlinear transformation (a broadly tuned texture grabber with a preference for low spatial frequencies) followed by standard motion analysis.

[3]Chubb, C. & Sperling, G. (1991). Texture quilts: Basic tools for studying motion-from-texture. *J. Math. Psychol. 35.* *Supported by AFOSR Visual Information Processing Program, Grant 91-0178.

Shih,shui-I and George Sperling. Cluster analysis as a tool to discover covert strategies. *Proceedings of the Eastern Psychological Association*, 1992, **63**, 41.

CLUSTER ANALYSIS AS A TOOL TO DISCOVER COVERT STRATEGIES.

Shui-I Shih and George Sperling. New York University.

Cluster analysis is proposed as a tool to discover whether and how performance in a series of sessions (or trials) depends on the effect of covert strategies. We illustrate its value by applying it to a large short-term memory experiment to analyze the change in individual performances with practice. Data were collected from 34+ sessions for two subjects. The results for each session were characterized by 54 dependent variables. ANOVA showed a main effect for *dependent variables* but no main effect for *sessions.* A Tukey test yielded highly significant interaction of *sessions* × *dependent variables* for both subjects, indicating that there were complex, essentially unanalyzable, changes in performance over sessions. On the other hand, cluster analysis discovered a partition of the sessions into two clusters with distinctively different performances for one of the subjects. The nearly chronological correlation between the clusters and sessions means that *practice* produced the change in strategy; strategy being defined by the performances in the clusters. In conclusion, we recommend using both a statistical model (ANOVA) and an analytical tool (cluster analysis) to discover and characterize the effects of practice and covert strategies.

George Sperling. Selective Attention to an Item Is Stored as a Feature of the Item. *Bulletin of the Psychonomic Society*, 1991, **29**, 473.

8:00-8:20 (1)

Selective Attention to an Item Is Stored as a Feature of the Item. GEORGE SPERLING, *New York University*, & STEPHEN A. WURST, *SUNY at Oswego*—Subjects must detect a repetition in a stream of 30 characters flashed at 10 per second. Items alternate in either color (black/white), size, orientation, or spatial frequency. Selectively attending a feature (e.g., black) never improves detection of repeated attended (black) versus unattended (white) items. Many counterintuitive results are explained by assuming (1) all items are stored in short-term memory (there is no perceptual filtering) and (2) attention to an item is itself stored as a feature of that item.

Anne Sutter, George Sperling and Charles Chubb. Furthur Measurements of the Spatial Frequency Selectivity of Second-Order Texture Mechanisms. Investigative Opthalmology and Visual Science, 1991, **32, No. 4**, *ARVO Supplement*, 1039

FURTHER MEASUREMENTS OF THE SPATIAL FREQUENCY SELECTIVITY OF SECOND-ORDER TEXTURE MECHANISMS

Anne Sutter, George Sperling, & Charles Chubb

Human Information Processing Laboratory, New York University, NY, NY 10003

A number of investigations of texture and motion perception suggest a two-stage processing system consisting of an initial stage of selective linear filtering, followed by a rectification and a second stage of selective linear filtering. Here we present new data measuring two properties of the second-stage filters: their contrast modulation sensitivity as a function of spatial frequency (MTF), and the relation of initial spatial filtering to second-stage selectivity. To determine the MTF, we used a staircase procedure to obtain amplitude modulation thresholds for the detection of the orientation of Gabor modulations of a bandlimited noise carrier. We used improved noise carriers with a narrower bandwidth than the stimuli reported last year. Four carrier bands were created with center frequencies of 2, 4, 8, and 16 c/deg. The spatial frequency of the test signals (Gabor amplitude modulations) ranged from 0.5 to 8 c/deg.

The improvements in our stimuli produced a different pattern of results: (1) The threshold amplitude of signal modulation was lowest for 0.5 and 1.0 c/deg. Above 1.0 c/deg, threshold increased with frequency[1]. (2) There was a significant interaction of carrier frequency band with the modulating frequency, with the lowest thresholds occuring for carrier frequency/modulation frequency ratios of about three to four octaves. These results indicate that the second-stage selective filters and detectors are most sensitive to frequencies lower than or equal to 1 c/deg, and that they are selective with regard to the spatial frequency content of the carrier noise on which the signals are impressed.

[1] Jamar, J.H.T. & Koenderink, J.J., (1985). *Vis. Res.* 25 (4) pp. 511-521.

Supported by AFOSR Life Sciences Directorate Grant 88-0140 and NIMH Grant 5T32MH14267.

Peter Werkhoven, Charles Chubb and George Sperling. Texture-Defined Motion is Ruled by an Activity Metric--Not by Similarity. Investigative Opthalmology and Visual Science, 1991, **32**, No. **4**, *ARVO Supplement*, 829

TEXTURE-DEFINED MOTION IS RULED BY AN ACTIVITY METRIC - NOT BY SIMILARITY

Peter Werkhoven, *Charles Chubb* and *George Sperling*.

Human Information Processing Laboratory, New York University

We examined motion carried by textural properties. The stimuli we used consisted of patches of sinusoidal grating of various spatial frequencies and contrasts. Phases were randomized to insure that motion mechanisms sensitive to correspondences in stimulus luminance were not systematically engaged.

We used an ambiguous apparent motion paradigm in which a "heterogeneous" motion path (defined by alternating patches of a type A and a type B texture) competes with a "homogeneous" motion path defined by patches of type A. We found that the strength of these (2nd order) motion stimuli is determined by the covariance of the *activity* of the textures that define the motion paths. The activity of a texture is an hypothesized property that is proportional to the texture's contrast and is found to be inversely proportional to its spatial frequency (within the range of spatial frequencies examined). Indeed, heterogeneous motion between equal contrast patches of a high spatial-frequency texture A and a low-spatial frequency texture B can easily dominate homogeneous motion between two patches of A because the activity of texture B is higher than that of texture A.

At temporal frequencies higher than 4 Hz, we find that activity covariance almost exclusively determines motion strength. At lower temporal frequencies, similarity between textures becomes a significant factor as well.

Supported by AFOSR Life Sciences, Visual Information Processing Program, Grant 88-0140.

Joshua A. Solomon and George Sperling. Can We See 2nd-Order Motion and Texture in the Periphery? Investigative Opthalmology and Visual Science, 1991, **32, No. 4,** *ARVO Supplement,* 714

CAN WE SEE 2nd-ORDER MOTION AND TEXTURE IN THE PERIPHERY?

Joshua A. Solomon and *George Sperling*,

Human Information Processing Laboratory, New York University

Stimuli. Our 1st-order stimuli are moving sine gratings. Our 2nd-order stimuli are patches of static visual noise, whose contrasts are modulated by moving sine gratings. Neither the spatial orientation nor the direction of motion of these 2nd-order (drift-balanced) stimuli can be detected by analysis of their Fourier domain power spectra. They are invisible to Reichardt and motion-energy detectors.

Method. For these dynamic stimuli, in the fovea, and at 12 deg eccentricity, we measured contrast modulation thresholds as a function of spatial frequency for discrimination of ±45 deg texture slant and for discrimination of direction of motion. Spatial frequency was varied by changing viewing distance.

Results. For sufficiently low spatial frequencies and sufficiently large contrast modulations, all stimuli are visible both foveally and peripherally. For peripherally viewed 1st-order gratings, the highest spatial frequency at which motion or texture discrimination is possible is about 1/4 that at which the corresponding discrimination is possible for foveally viewed gratings. For peripherally viewed 2nd-order gratings, the highest spatial frequencies at which motion or texture discrimination are possible are somewhat less than 1/4 the frequencies of the corresponding foveal discriminations. Thus, as the stimulus moves peripherally, the visual mechanisms that detect 2nd-order motion and texture lose sensitivity somewhat faster than the 1st-order mechanisms.

Conclusions. Under certain specific assumptions, our results suggest the following about the neural detectors involved in these discriminations: (1) For both motion and texture, there are more foveal than peripheral detectors at all spatial frequencies. (2) There are more 1st-order than 2nd-order detectors. (3) On the average, foveal detectors respond to higher spatial frequencies than peripheral detectors. (4) The 2nd-order foveal–peripheral spatial frequency difference is somewhat larger than the 1st-order difference.

Supported by AFOSR Life Sciences, Visual Information Processing Program, Grant 88-0140.

Vision Res. Vol. 33, No. 18, pp. 2671–2683, 1993
Printed in Great Britain. All rights reserved

0042-6989/93 $6.00 + 0.00

The Lateral Inhibition of Perceived Contrast is Indifferent to On-Center/Off-Center Segregation, but Specific to Orientation

JOSHUA A. SOLOMON,[*] GEORGE SPERLING,[†] CHARLES CHUBB[‡]

Received 27 January 1993; in revised form 21 May 1993

When a central test patch C, composed of an isotropic spatial texture, is surrounded by a texture field S, the perceived contrast of C depends substantially on the contrast of the surround S. When C is surrounded by a high contrast texture with a similar spatial frequency content, it *appears* to have less contrast than when it is surrounded by a uniform field. Here, we employ two novel textures: T^+ which is designed to selectively stimulate only the on-center system, and T^-, the off-center system. When C and S are of type T^+ and T^-, the reduction of C's apparent contrast does not vary with the combination of T^+, T^-. This demonstrates that the reduction of C's apparent contrast is mediated by a mechanism whose neural locus is central to the interaction between on-center and off-center visual systems. We further demonstrate *orientation specificity*: the reduction of grating C's apparent contrast by a su. ;ound grating S, of the same spatial frequency is greatest when C and S have equal orientation. Using dynamically phase-shifting sinusoidal gratings of 3.3, 10 and 20 c/deg, we measured reduction of apparent contrast using different contrast-combinations of C and S. Results: (1) S gratings, both parallel and perpendicular to C, cause a reduction in C's apparent contrast relative to a uniform surround. (2) In all of the viewing conditions, the reduction of apparent contrast induced by the parallel surrounds was at least as great as that induced by the perpendicular surrounds. Often it was much greater. (3) Orientation specificity increases with increasing spatial frequency and with decreasing stimulus contrast.

Lateral inhibition Orientation specificity Contrast perception Texture Scale invariance

INTRODUCTION

Previously, we demonstrated that the perceived contrast of a patch of isotropic, random visual texture is diminished when that patch is embedded in a surrounding background of similar texture (Chubb, Sperling & Solomon, 1989). We also demonstrated that, for brief flashes of the center and surround, this contrast inhibition effect is strictly monocular. That is, when the patch and the surrounding texture are presented to different eyes, the apparent contrast of the center will not be diminished. In addition, we showed that this effect is spatial-frequency specific: when the spatial frequency of the patch differs by an octave from the frequency of the surround, then the apparent contrast of the patch is influenced very little by the contrast of the surround. These results suggest the existence, at some level of visual processing, of laterally-interactive neural arrays tuned to local contrast energy within relatively narrow spatial frequency bands. Neural arrays of this type have also been suggested by other psychophysical and physiological studies (Chubb & Sperling, 1988, 1989; Shapley & Victor, 1978; Enroth-Cugell & Jakiela, 1980; Ohzawa, Sclar & Freeman, 1985; Sagi & Hochstein, 1985; Heeger, 1992).

The present research describes two new phenomena of lateral texture–contrast interactions. The first section (Expt 1) demonstrates that signals from on-center and off-center visual mechanisms are combined prior to processing by the mechanism which mediates the lateral inhibition of perceived contrast. The second section (Expts 2–4) demonstrates that the neural arrays which compose this laterally interactive mechanism are tuned to specific orientations of spatial texture, and measures the orientation specificity as a function of the contrasts of the center and surround.

GENERAL METHODS

Subjects

In each experiment two subjects were run. Each subject was a trained psychophysical observer (JS and

[*] Syracuse University, Institute for Sensory Research, Merrill Lane, Syracuse, NY 13244-5290, U.S.A.
[†] To whom all correspondence should be addressed at: Department of Cognitive Science, University of California, Irvine, CA 92717, U.S.A.
[‡] Psychology Department, Rutgers University, Busch Campus, New Brunswick, NJ 08903, U.S.A.

CC are experimenters). Each had normal or well-corrected vision.

Stimuli

Each stimulus consisted of a circular patch of texture (the *center*) surrounded by another circular patch of texture (the *surround*). The mean luminance of each center and surround was the same, and equal to the background of the display. All displays were presented at 60 frames/sec, and all stimuli were dynamic. That is, new random phases of the textures in the center and in the surround were selected every $\frac{1}{15}$ sec. The images were created using both specially designed programs and the HIPS image-processing software package (Landy, Cohen & Sperling, 1984).

Apparati

The displays for the experiments were presented on three different monochrome graphics monitors using an Adage RDS 3000 image display system. In Expt 1, subject CC used a Leading Technologies 1230V (12 in diagonal) with a mean luminance of 90 cd/m^2, and subject JS used a US Pixel PX-15 (15 in diagonal) with a mean luminance of 40 cd/m^2. In Expts 2 and 3, both subjects used a Princeton MAX-15 (14 in diagonal), with a mean luminance of roughly 60 cd/m^2. In Expt 4, both subjects used the US Pixel.

Calibration

For each monitor, luminance linearization was achieved using a center/surround display comprised of a uniform circular patch surrounded by an annular background containing a squarewave pattern of spatial frequency equal to that of the sinusoidal pattern used in Expts 2–4. A sheet of frosted plastic was placed in front of the monitor. At distances of 1 m or more, this effectively filtered out the high spatial frequencies in the annular surround, and both center and surround appeared uniform. The experimenter set the maximum and minimum luminance values for the light and dark pixels of the surround, and then adjusted the luminance of the center until center and surround were no longer distinguishable. The resulting center luminance is thus halfway between the maximum and minimum luminances of the display. Systematic iterations of this technique yield displays with precisely calibrated contrasts. In order to stabilize the monitor's power draw throughout the linearization process, two separate center/surround displays were shown concurrently. When establishing a relatively high luminance value on one display, the corresponding low luminance value was established on the other.

Procedure

The subject sat in a dark room and viewed the display binocularly. The only source of illumination was the light from the continuously illuminated display. The trial sequence is illustrated in Fig. 1. Upon a key press, a stimulus with a center and a surround was presented. Then, the central texture was presented alone, then the center/surround pair again and finally just the center again. Each of the four presentations lasted 533 msec. The subject's task was to make a forced-choice judgment. The subject had to decide whether the central grating had more contrast in the presence of the surround or when it appeared alone. The subject indicated his/her choice by pressing one of two buttons. There was no limit to the time within which the subject had to give his/her answer. In summary: the center appeared four times in a trial, twice with the surround on ("masked center"), and twice with the surround off ("test center"). The subject's task was to decide whether the apparent contrast of the test center was greater or less than the apparent contrast of the masked center.

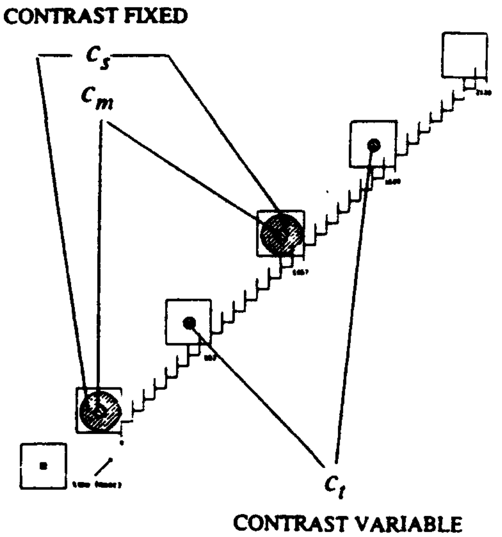


FIGURE 1. Illustration of general procedure. Subject fixates on a cue spot. Following a key press, eight frameblocks (four frames at 60 frames/sec) appear in one of one of the four center/surround texture combinations. They are followed by eight frameblocks of just the center texture (surround contrast equals zero). This 16 frameblock sequence is then repeated. Presentation rate is 15 frameblocks sec. Immediately following the sequence a blank frame is presented, which is terminated by the subject's response. The subject's task is to indicate whether or not the center texture appeared to have more contrast in the presence of the surround than when viewed in isolation.

We use "$c_t \hat{>} c_m$" to denote a response indicating that the apparent contrast of the test center was greater than apparent contrast of the masked center and "$c_t \hat{<} c_m$" to denote the response that the apparent contrast of the test center was less than apparent contrast of the masked center. Consider the psychometric function mapping c_t, the actual contrast of the test center, to P("$c_t \hat{>} c_m$"). We determined two points on this function, the values of c_t for which P("$c_t > c_m$") = 0.62 and 0.38. This allows us to estimate both the point of subjective [the value of c_t for which P("$c_t \hat{>} c_m$") = 0.5] and the slope of the psychometric function, which is a measure of the intrinsic variability of the point of subject of subjective equality. To determine these points, we used a staircase procedure in which the subject's response on trial n is used to determine the contrast of the test center on trial $n + 1$.

For each stimulus, there were two interleaved staircases, designated by their expected points of convergence on the psychometric function: a 0.62 staircase and a 0.38 staircase. In the 0.62 staircase, the contrast of the test center was decreased by one step size, after every "$c_t \hat{>} c_m$" response. The contrast of the test center was increased by one step size, after two consecutive trials yielding "$c_t \hat{<} c_m$" responses. In the 0.38 staircase, the contrast of the test center was decreased by one step size, after every "$c_t \hat{<} c_m$" response. The contrast of the test center was increased by one step size, after two consecutive trials yielding "$c_t \hat{>} c_m$" responses.

Specifically, we measured the reduction of the masked center's apparent contrast induced by the presence of the surround, as a percentage. The calculation of this is shown here

$$\text{percent reduction in apparent contrast} = 100\left[\frac{c_m - c_t}{c_m}\right] \quad (1)$$

where c_m is the actual contrast of the masked center, and c_t is the actual contrast of the test center.

For each viewing condition in each experiment, subjects ran one block of 50 trials (at least six trials per staircase) using a step size of value $\frac{1}{10}c_t$. Then, with a smaller step size (approximately $\frac{1}{40}c_t$), subjects ran as many blocks of 100 trials as necessary (typically 3–5) until the variance of the reversed points of each staircase, divided by the square root of the number of reversals, was no greater than 2.5%.

EXPERIMENT 1: ON-CENTER/OFF-CENTER INTERACTION

The fact that Chubb *et al.* (1989) observed the induced reduction of apparent contrast to be strictly monocular in their conditions suggests it is a relatively low level visual process. This raises the possibility that the lateral inhibition underlying the effect might be occurring at the level of on-center and off-center retinal ganglion cells or LGN cells. If so, it seems possible that the inhibition would selectively occur between cells of the same contrast polarity. In other words, perhaps on-center cells selectively inhibit other on-center cells and off-center cells selectively inhibit other off-center cells. Experiment 1 investigates this conjecture.

Stimuli

The on-center and off-center visual pathways work in tandem to efficiently code information about contrast in the visual field. Both on-center and off-center ganglion and LGN cells maintain a steady base rate of firing, which can be increased or decreased by appropriate stimuli. It seems unlikely that contrast information from suprathreshold stimuli can be adequately signaled by *decreases* in the base firing rates. In the extreme, no cell can distinguish between two stimuli, each of which has sufficient contrast to cause a complete cessation in firing. Less extreme stimuli may slow the firing rate down enough so that the rate itself may only become discernible to subsequent processing stages after some considerable time (Enroth-Cugell & Robson, 1984). However, stimuli of contrast which cause a decrease in the firing rate of on-center cells should simultaneously cause an increase in the firing rate of off-center cells, and vice versa. Thus, contrast information can be adequately coded by an increase in the firing rate of one of the two systems.

Indeed, selective, pharmacological blocking of on-center cells in monkeys has been demonstrated (Schiller, Sandell & Maunsell, 1986) to severely impair detection of bright spots, without affecting dark spot detection. This finding supports the notion that local luminance increments are coded by the on-center system, and local luminance decrements are coded by the off-center system.

Two recent psychophysical studies with human subjects supply further evidence for segregated processing of local luminance increments and decrements. Malik and Perona (1990) demonstrated that when one texture is defined by patches composed of light bars with dark sidebands, and another by dark bars with bright sidebands, a boundary between the two textures is perceived preattentively. Solomon and Sperling (1993) demonstrated that one-third of the population can perceive the motion of gratings defined by the same textures used in the current experiment. A mechanism having a linear function of stimulus luminance as input would not be able to segregate Malik and Perona's textures nor extract motion from Solomon and Sperling's gratings. Neither would one whose input equally weights local luminance increments and decrements. However, performance of these tasks *can* be modeled by a mechanism whose input effectively filters out either local luminance increments or local luminance decrements, and has a soft activation threshold.

Based on the luminance-balanced micro-elements of Carlson, Anderson and Moeller (1980), two novel textures were designed to investigate mechanisms which receive input from either on- or off-center neurons, but not both. These textures consist of bright or dark points on gray backgrounds. In theory, bright points will selectively increase the firing rates of on-center cells in whose receptive field centers they fall, and dark points will increase the firing rates of off-center cells in whose receptive field centers *they* fall. These textures are somewhat similar to the stimuli used by Zemon, Gordon and Welch (1988), in an attempt to differentially stimulate the on- and off-center systems. Ours differ from the textures used by Zemon *et al.*, in that ours are designed so that the level of adaptation of neurons in each pathway remains constant, independent of the polarity of the texture. This is accomplished by ensuring that the mean luminance of all textures remains constant and that phase (i.e. the positions of the bright and dark points) is randomly determined every $\frac{1}{15}$ sec. Unlike the static textures used by Zemon *et al.* which were not equated for mean luminance, our textures are designed so that any neuron with a receptive field large enough to include several bright or dark points will receive the same stimulation.

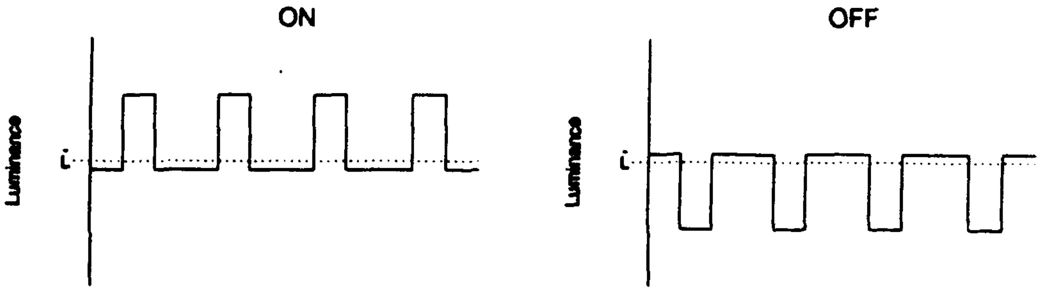


FIGURE 2. Illustration of "ON" and "OFF" textures. A vertical (or horizontal) slice through each texture is diagrammed. Mean luminance $\bar{L}$, is indicated on the ordinates.

Insofar as the on-center and off-center ganglion cells can be modeled as having center–surround antagonism and a (soft) threshold for firing, then the bright spots in our textures will selectively increase the firing rates of on-center cells in whose receptive field centers they fall, and the dark spots will increase the firing rates of off-center cells. Various plausible assumptions about the responsiveness of on- and off-center systems make a high

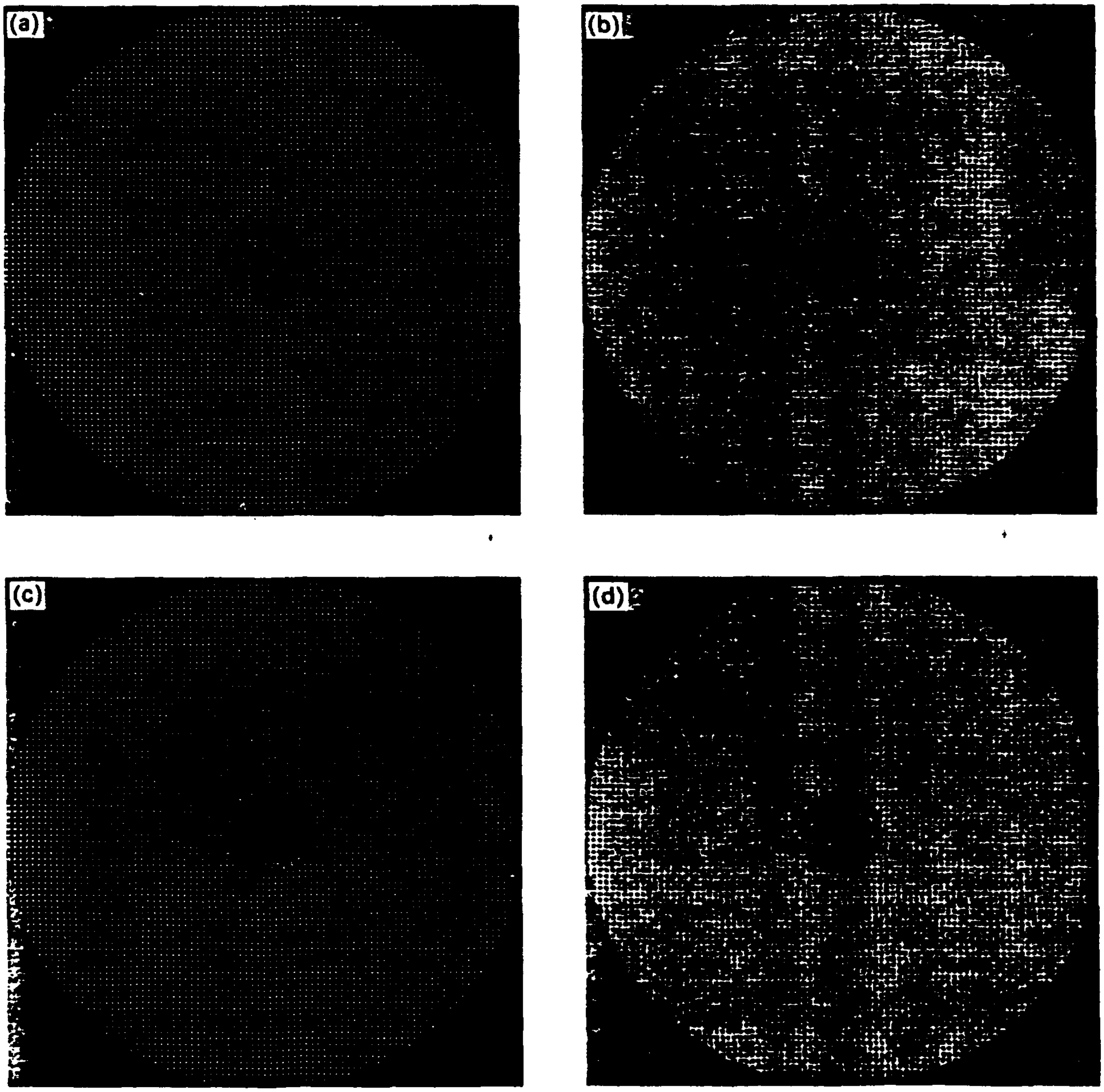


FIGURE 3. Stimuli for Expt 1. (a) On-center stimulating center, on-center stimulating surround (ON/ON). (b) On-center stimulating center, off-center stimulating surround (ON/OFF). (c) Off-center stimulating center, on-center stimulating surround (OFF/ON). (d) Off-center stimulating center, off-center stimulating surround (OFF/OFF).

*See the attached sheet for better figure.

degree of physiological selectivity likely. For example, if responses are proportional to the second power of contrast of near-threshold stimuli, then stimulation of the on-center system by bright points should dwarf any concomitant stimulation of the off-center system by their background. Nonetheless, the true physiological selectivity of these textures remains to be tested.

The texture designed to selectively stimulate on-center cells is comprised of a regular grid of bright pixels (the pixel at every third row and every third column is bright). This texture is called an "ON" texture. The "OFF" texture is designed to selectively stimulate off-center cells; it is comprised of a regular grid of dark pixels (the pixel at every third row and every third column is dark). The luminances of the other pixels in the textures are chosen so that the mean luminance of the ON texture is equal to the mean luminance of the OFF texture (see Fig. 2).

The stimuli used in this experiment were composed of center/surround combinations of these textures. The positions of the pixel grids in each center and each surround of each block of four frames ($\frac{1}{15}$ sec) were randomly chosen from one of nine possible phases (three horizontal positions times three vertical positions); this produces a dynamically changing display and the appearance of a jittering boundary between center and surround.

There were four stimulus combinations corresponding to two different types of surround texture times two different types of center texture. The four center/surround combinations are shown in Fig. 3. ON *masked* centers (with a surround) are judged only relative to ON *test* centers (without a surround), and OFF test centers are judged only relative to OFF masked centers.

The stimuli were viewed from a distance of 0.67 m. At this distance, for JS the surround subtended a visual angle of 9.3 deg, and the center, 1.5 deg. For CC the surround subtended a visual angle of 7.2 deg, and the center, 1.2 deg.

Results and discussion

The results for both subjects are plotted in Fig. 4. Two points, the means of the staircases with the different convergence points, are shown for each center/surround combination. The lower point indicates the percent reduction in apparent contrast, as determined by the 0.38 staircase; the upper point indicates the 0.62 staircase. Symbol size reflects maximum standard error. Most standard errors are much less than symbol size.

For each center/surround combination, both subjects show more than a 50% reduction of the center's apparent contrast induced by the surround. The mean percent reduction (mean of 0.62 and 0.38 staircases) of apparent contrast does not vary with center/surround combination. A surround which is intended to excite only the off-center visual system causes the same degree of reduction in the apparent contrast of a center which is intended to excite only the on-center visual system, as does a surround which is intended to excite only the on-center system, and vice versa.

According to these results, the neural mechanism that mediates the lateral interactions responsible for this reduction of apparent contrast combines information from both the on-center and the off-center pathways. The mechanism for the lateral inhibition of perceived contrast lies central to the point of on-center/off-center integration.

EXPERIMENTS 2-4: ORIENTATION SPECIFICITY

The procedure we use here was motivated by the initial observation that a surround grating whose overall contrast is temporally modulated will cause an apparent, opposite phase modulation in the contrast of a temporally constant target grating. When two target gratings are used, one with orientation parallel to that of the surround and one with orientation perpendicular to that of the surround, the contrast of the parallel target seems to modulate more than the contrast of the perpendicular target. Further observations suggested that the disparity between contrast inhibition induced by parallel and perpendicular surround gratings was not always pronounced. Some stimulus parameters are better than others at eliciting orientation specific differences in contrast inhibition. The following

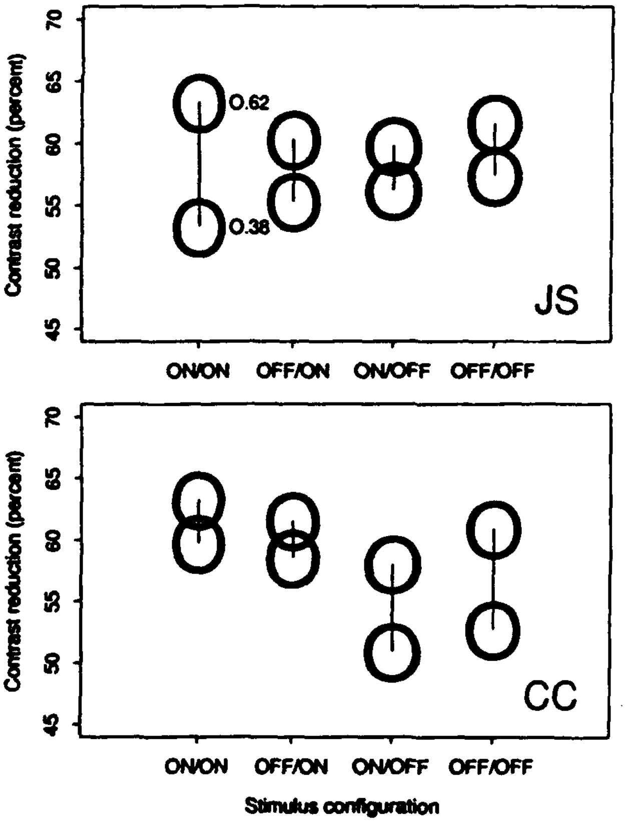


FIGURE 4. Results for subjects JS and CC, Expt 1. For each stimulus configuration, there is a 0.62 probability that the apparent reduction in contrast induced by the presence of the surround is less than that denoted by the upper point. Likewise, there is a 0.38 probability that the apparent reduction in contrast induced by the presence of the surround is less than that denoted by the lower point. Symbol size reflects maximum standard error.

experiments were designed to confirm our initial observations.

Stimuli

All the stimuli were center/surround combinations of sinewave gratings. For each $\frac{1}{15}$ sec frameblock of the stimulus, the phases of both the center and the surround gratings were independent and randomly determined at one of four possible phases. The sinewave gratings were presented in one of two different orientations: either slanted 45 deg in one direction or slanted 45 deg in the other direction. There were four center/surround combinations, corresponding to the two different orientations of surround grating times the two different orientations of center grating. The four stimulus combinations are illustrated in Fig. 5.

Procedures

There were four independent variables: center/surround orientation (parallel, perpendicular), spatial frequency (3.3, 10, 20 c/deg), contrast of the surround c_s, and contrast of the masked center c_m. Spatial frequency was varied by varying viewing distance; this had the virtue of leaving all the physical characteristics of the display intact and varying only the retinal scale. The dependent variable was the *percent reduction in apparent contrast* of the center induced by the presence of the surround, as defined in equation (1). Viewing conditions and results for Expts 2–4 are summarized in Table 1.

The monitor used to display the stimuli in Expt 4 was different from the one used to display the stimuli in Expts 2 and 3. As a consistency check, both subjects performed the 3.3 c/deg, $c_s = 1.0$, $c_m = 0.5$ viewing condition with the new monitor. The resulting data were indistinguishable from the initial data gathered in Expt 2.

Only with the $c_s = 1.0$, $c_m = 0.5$ procedure was the center grating visible at 4 m. (At this, the longest viewing distance, the center grating had a spatial frequency of 20.0 c/deg.) Thus, this viewing distance was omitted from all other procedures. Similarly, with the $c_s = 0.04$, $c_m = 0.03$ procedure, the center grating was invisible from 2 m. Thus, only the shortest viewing distance was used in Expt 4.

Results

There were no systematic differences between the responses to stimuli of reflectively symmetrical orientations; therefore, these data have been pooled. That is, the data from trials in which the center and surround shared the same orientation have been pooled (*parallel configuration*), and the data from trials in which the center and surround were perpendicularly oriented have been pooled (*perpendicular configuration*).

The results for Expts 2–4 are plotted in Figs 5 and 6. As in Expt 1, two points are plotted for each configuration to indicate the points of convergence of the 0.62 staircases and the 0.38 staircases.

Each individual graph compares the reduction in apparent contrast for the parallel stimulus configuration with that for the perpendicular stimulus configuration.

Trends in the data. (1) For every stimulus configuration, in every viewing condition, there is a statistically significant ($P < 0.005$) percent reduction of the center's apparent contrast induced by the surround.

(2) The difference between the heights of the parallel-configuration points and the perpendicular-

TABLE 1. Viewing conditions and results for Expts 2–4

								Contrast reduction (%)			
		Contrast		dva		Spatial frequency (c/deg)	Mean luminance (cd/m^2)	Parallel		Perpendicular	
Experiment	Subject	c_s	c_m	Surround	Center			0.707	0.293	0.707	0.293
2	AS	1.0	0.5	9.9	1.64	3.3	60	47.2	41.9	44.6	40.6
2	AS	1.0	0.5	3.3	0.55	10.0	60	51.5	45.4	42.9	34.0
2	AS	1.0	0.5	1.6	0.28	20.0	60	52.9	37.1	21.7	6.0
2	AS	0.2	0.1	9.9	1.64	3.3	60	38.1	28.5	30.0	22.2
2	AS	0.2	0.1	3.3	0.55	10.0	60	52.6	36.7	17.4	4.4
2	JS	1.0	0.5	9.9	1.64	3.3	60	29.3	19.7	27.5	19.0
2	JS	1.0	0.5	3.3	0.55	10.0	60	37.7	27.4	28.5	18.1
2	JS	1.0	0.5	1.6	0.28	20.0	60	54.2	28.7	22.9	4.6
2	JS	0.2	0.1	9.9	1.64	3.3	60	26.3	19.3	23.0	8.1
2	JS	0.2	0.1	3.3	0.55	10.0	60	41.4	29.3	19.6	7.7
3	AS	1.0	0.1	9.9	1.64	3.3	60	49.6	41.4	40.0	31.5
3	AS	1.0	0.1	3.3	0.55	10.0	60	58.1	47.8	37.0	24.4
3	AS	0.2	0.1	9.9	1.64	3.3	60	35.6	27.4	29.6	24.4
3	AS	0.2	0.1	3.3	0.55	10.0	60	52.6	43.7	26.3	17.0
3	JS	1.0	0.1	9.9	1.64	3.3	60	44.4	34.8	42.6	29.3
3	JS	1.0	0.1	3.3	0.55	10.0	60	41.1	30.4	28.5	4.4
3	JS	0.2	0.1	9.9	1.64	3.3	60	26.6	19.3	23.0	8.1
3	JS	0.2	0.1	3.3	0.55	10.0	60	41.1	29.3	19.6	7.7
4	AS	1.0	0.04	9.9	1.64	3.3	40	64.6	60.9	62.0	56.9
4	AS	0.04	0.03	9.9	1.64	3.3	40	37.8	31.3	27.8	12.8
4	JS	1.0	0.04	9.9	1.64	3.3	40	48.5	43.0	42.5	28.5
4	JS	0.04	0.03	9.9	1.64	3.3	40	40.5	24.4	23.6	6.6

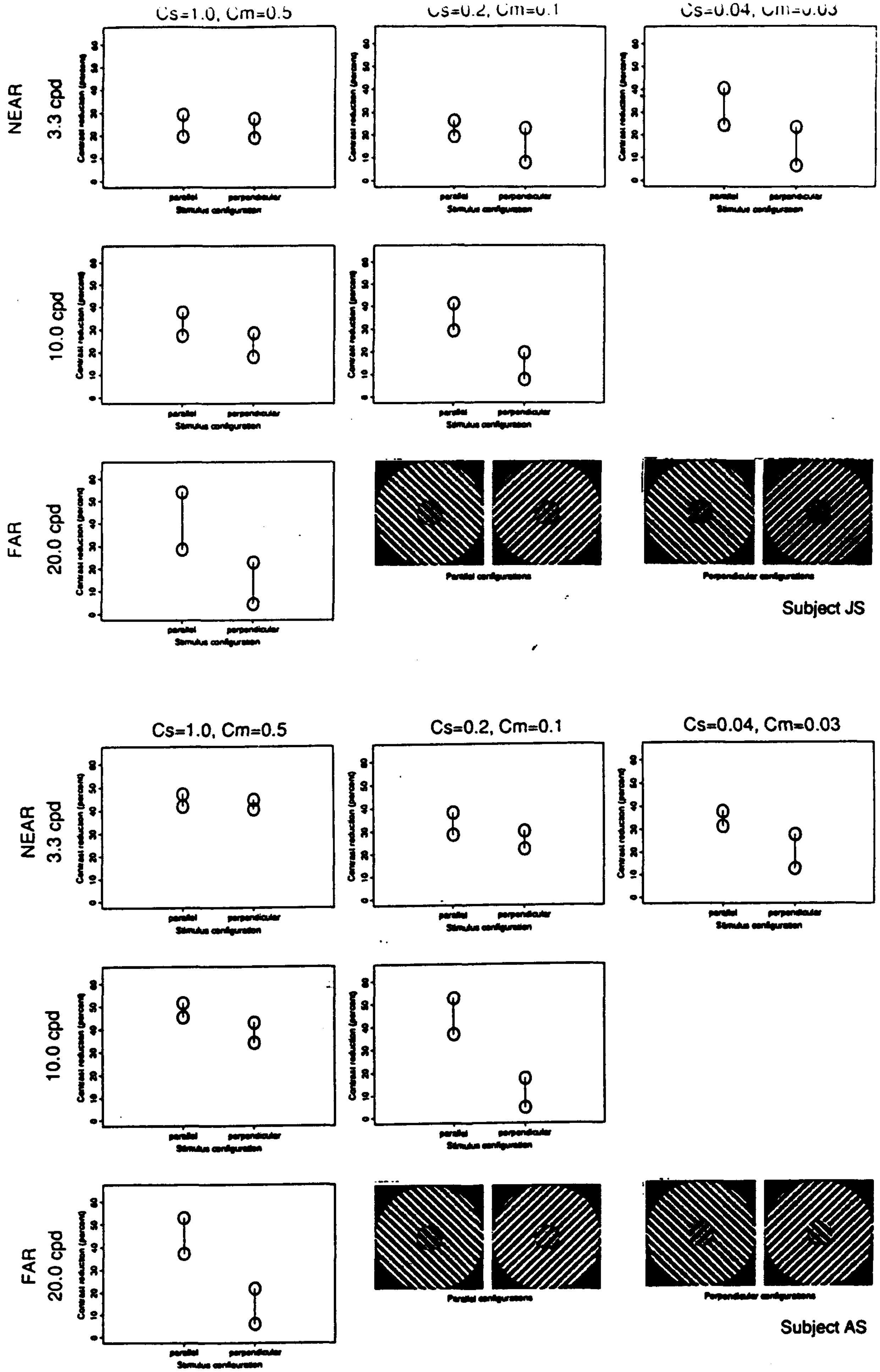


FIGURE 5. Results for subjects JS and AS, Expts 2 and 4. Six experimental conditions, four types of stimulus (see insets). Data from reflectively identical center/surround combinations are collapsed.

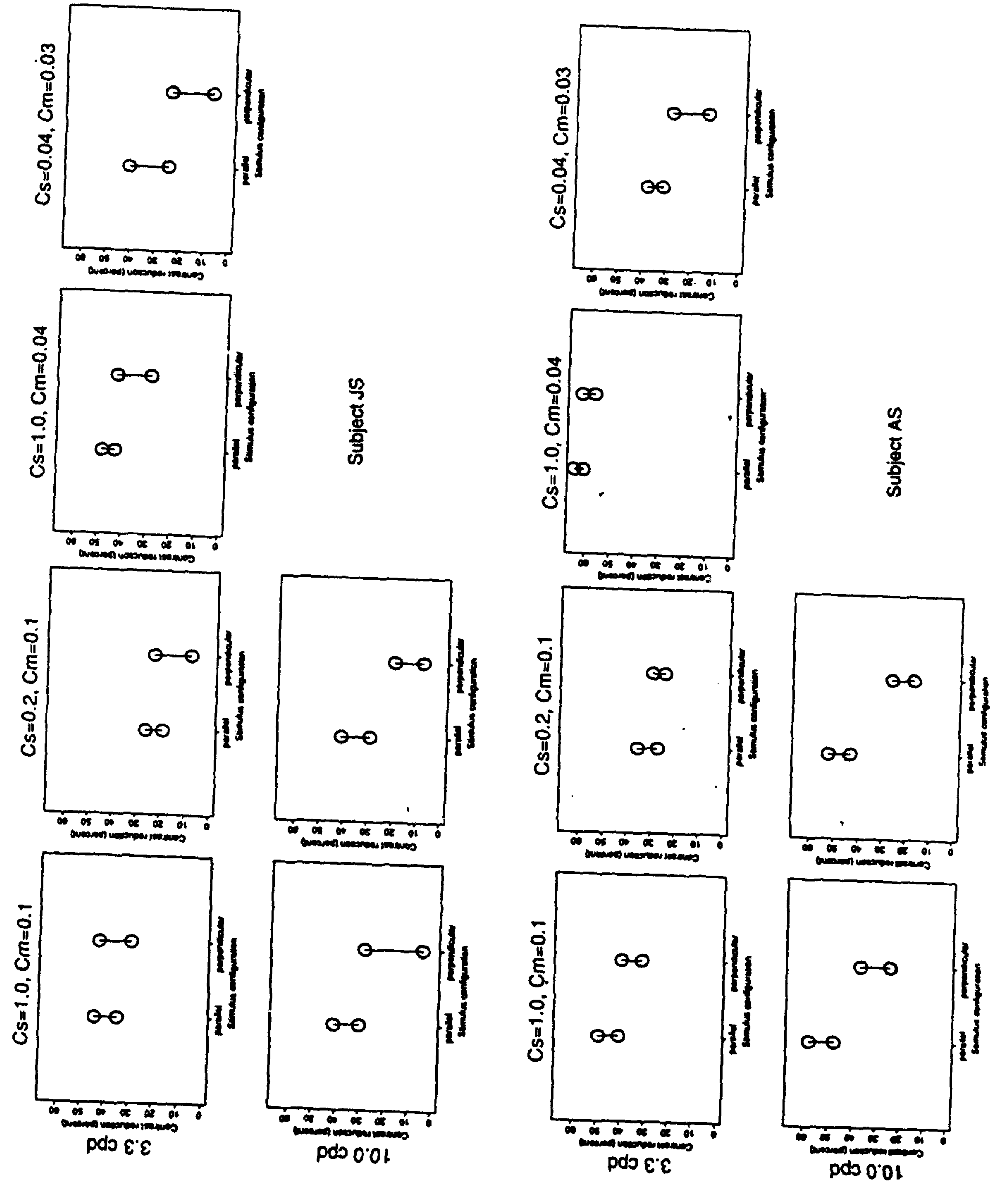


FIGURE 6. Results for subjects JS and AS, Expt. 3.

configuration points on each graph is a measure of the orientation specificity in that viewing condition. In all of the viewing conditions, the percent reduction of apparent contrast induced by the parallel surrounds is at least as great as that induced by the perpendicular surrounds. Often it is much greater.

(3) The left column of Fig. 5 represents all of the data from the trials in which $c_s = 1.0$, $c_m = 0.5$. Note that *an increase in the viewing distance (and hence an increase in the retinal spatial frequency of the gratings) results in greater orientation specificity*. This general trend obtains for the other combinations of c_s and c_m (middle column of Fig. 5, two leftmost columns of Fig. 6).

(4) The first and fourth rows of Fig. 5 represent data from trials in which the retinal spatial frequency of the gratings was 3.3 c/deg, and the ratio of c_s:c_m was 2:1. Note that *a decrease in stimulus contrast results in an increase in orientation specificity*. This general trend also holds for the 10.0 c/deg stimuli (second and fifth rows of Fig. 5).

GENERAL DISCUSSION

Chubb *et al.* (1989) conducted similar experiments to those reported here. However, they used patches of isotropically filtered visual noise rather than sinusoidal gratings. Their principal findings were: (i) for high contrast surrounds, when c_m was roughly equal to half the surround contrast, percent reduction of apparent contrast was around 40%, *provided* that center and surround were filtered into the same frequency band; (ii) if the center and surround were presented to opposite eyes, no induction occurred; (iii) if center and surround were filtered into octave-wide frequency bands, with center frequencies one octave apart, the percent reduction of apparent contrast dropped down to 15%. This third result indicates that the reduction of apparent contrast induced by the presence of the surround is spatial frequency specific.

The current experiments investigate the degree to which this reduction of apparent contrast induced by the surround is *orientation* specific.

Channels, tuned filters, neurons

Since the pioneering work of Campbell and Robson (1968), it has been recognized that the visual system filters the visual signal into a number of relatively narrow spatial frequency bands, which they termed *channels*. Each of these channels can be modeled approximately as an array of linear filters with all filters in the array sharing the same receptive field profile, but centered at different retinal locations so as to cover the visual field. Each of these filters produces a positive or negative output in response to any given stimulus. Apparent contrast is proportional to the absolute value of filter output.

One way of understanding the results of Chubb *et al.* (1989) is to suppose that the output values produced by the filters in these arrays are subject to lateral inhibition from other filters in the same array. In particular, the higher the *absolute value* of the output of a filter in such an array, the greater its inhibitory effect on other filters near it in the array. Thus, high contrast regions of a narrow band texture produce regions of high absolute value in the filter array tuned to that texture; in turn, these regions of high absolute value output act laterally to damp the magnitude of the output values produced by filters in nearby regions of the array, thereby lowering the apparent contrast of the inhibited region.

In the visual system, filters are realized by neurons. We assumed that, in each of our experimental conditions, the observed percent reduction of apparent contrast depends on the amount of lateral inhibition delivered to neurons tuned to the center texture by neurons tuned to the surround texture. For any viewing condition, the observed reduction of apparent contrast induced by a parallel surround is always at least as great as that induced by a perpendicular surround; we thus infer that the neurons tuned to the parallel surround deliver at least as much inhibition to the similarly tuned neurons being stimulated by the center texture than do the neurons tuned to the perpendicular surround. That is, neurons tuned to the same orientation deliver more inhibition to each other than do neurons tuned to different orientations.

Relations to physiology

Physiological studies of macaque and cat have yielded no evidence for any precortical orientation specificity (Hubel & Weisel, 1977). This restricts the neural locus of the interaction between texture-sensitive neurons. Equally restrictive is the result that surround-induced apparent contrast reduction is a strictly monocular effect.

When we first reported that the lateral inhibition of perceived contrast does not spread interocularly (Chubb *et al.*, 1989), we used tests involving only band-passed isotropic texture to support our claim. To insure that this result held true for high frequency gratings as well, we re-ran our "interocular induction" experiment with two subjects. In this procedure the center and surround, both 20 c/deg, were presented to different eyes in a continuous display. Here, and in the interleaved same-eye control trials, center and surround were separated by a thin gray annulus to prevent rivalry. The surround flashed either on or off every 500 msec. Subjects adjusted the contrast of the surround-on center, until it appeared equal to that of the surround-off center. As before, this manipulation was effective in removing any noticeable interaction between the contrast of the surround and the appearance of the center. Thus, we maintain that the neural locus for the lateral interaction between texture-sensitive neurons lies at an early cortical or precortical level of processing.

Physiological studies of the functional architecture of macaque and cat visual cortex have revealed that, outside of layer IV in area 17, binocularly driven cells greatly outnumber monocularly driven cells. Thus we propose that it is the neurons of this layer which combine texture information, in a spatially antagonistic

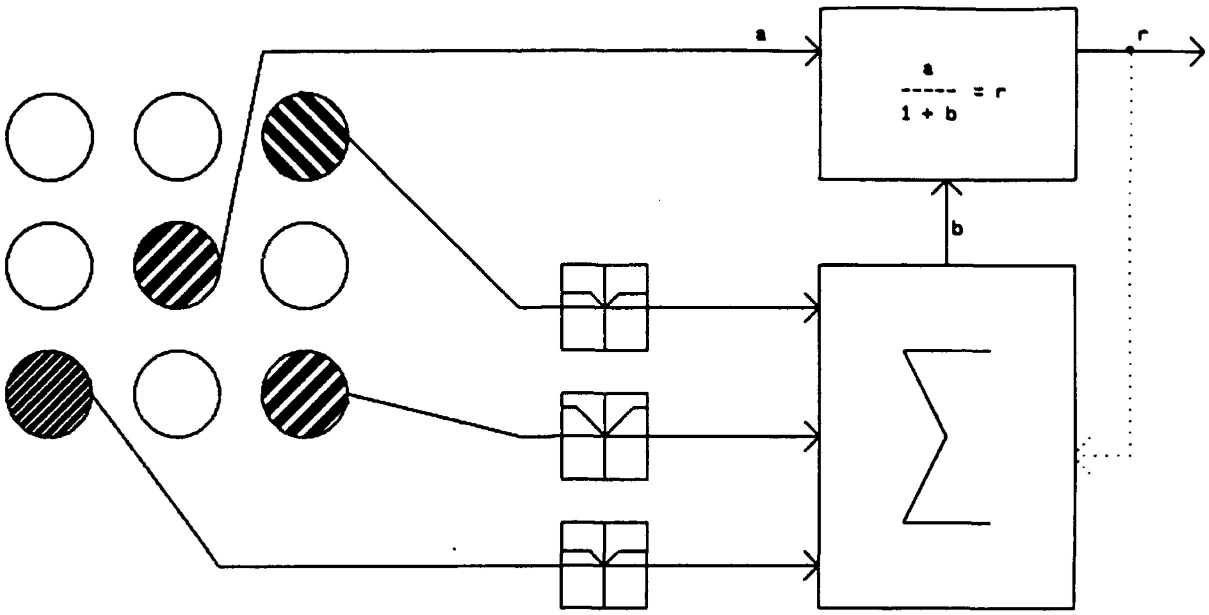


FIGURE 7. Model for the lateral inhibition of perceived contrast. Central unit tuned to specific spatial frequency and orientation. Excitatory component a, is the dot product (correlation) of the stimulus with the receptive field of the central unit. Surrounding units are tuned to a variety of frequencies and orientations. Their outputs are rectified and summed, giving preferential weighting (indicated by the filters in the small boxes) to those units with spatial location, frequency and orientation tuning similar to that of the central unit. The response r, of the central unit is scaled with respect to this combination b. The dotted arrow indicates a more complex model, in which the responses of surrounding units are scaled with respect to the response of the central unit.

way, resulting in surround-induced apparent contrast reduction.

A model

Simple theory: one-way interactions. Figure 7 diagrams proposed interactions between various texture-sensitive units. We use the term *units* (rather than *neurons*) because neurons transmit only positive firing rates (positive signals); it requires a push–pull pair of neurons (a neural *unit*) to transmit both positive and negative signals. Also, we do not differentiate here between a single neuron and many similar neurons that may be acting in concert.

The central unit is tuned to a specific spatial frequency and orientation. The excitatory component a is the dot product (correlation) of its receptive field with the stimulus. The surrounding units are tuned to a variety of spatial frequencies and orientations. Their outputs are first rectified (absolute value) and then added together giving preferential weighting (indicated by the filters diagrammed in the small boxes) to those units with spatial location, frequency and orientation tuning similar to that of the central unit. The output response r, of the central unit is scaled with respect to the rectified sum of surrounding outputs b. We consider this simple model first, and then a more complex model in which the interactions are reciprocal, the output of the surround units being scaled by the rectified output of the center unit.

Since virtually every viewing condition results in some reduction in apparent contrast, it is possible to construct a model that attributes a proportion p of this effect to a balanced mixture of parallel and perpendicularly oriented units that have precisely equivalent properties and occur in precisely the same numbers. Consequently, any orientation specific effect must be attributed to parallel and perpendicular units that have different properties and may occur in different numbers. Alternatively, one could attribute a proportion $q \leqslant p$ of the total observed reduction in apparent contrast to a population of unoriented receptive fields. For conceptual simplicity, for the proportion p of orientation-balanced units, we do not discriminate the balanced mixture of parallel and perpendicular receptive fields from a functionally equivalent mixture of unoriented receptive fields.

The model portrays inhibition as a divisive (shunting) form of gain control (Sperling & Sondhi, 1968; Sperling, 1970), for which *percent reduction in apparent contrast* is the natural dependent variable. The model is similar in spirit to the models proposed by Sperling (1989) and Heeger (1992). It differs in three respects: it deals in detail with the contrast saturation functions that limit lateral interactions, it allows for orientation specific normalization, and reciprocal inhibitory interactions between center and surround are treated explicitly.

To apply the simple model to the current experiments, we consider the equilibrium state when a masked center (contrast c_m) with its surround (contrast c_s) is equated in apparent contrast to the isolated test center (contrast c_t). Because the surround inhibits the masked center, the match is represented as

$$c_t = \frac{c_m}{1+b'}, \qquad b' = \sum_i w_{i,\theta} g_\theta(c_s). \qquad (2)$$

The functions g_θ are monotonically increasing functions that represent the influence of the surround on the center; $g_\theta = g_\parallel$ or $g_\perp$ depending on whether the orientation of the center is parallel (||) or perpendicular (⊥) to the surround. The values of the weights $w_{i,\theta} \geqslant 0$ depend on the relative orientations θ of the units, as well as their retinal locations i. Solving equation (2) for c_t and substituting for c_t in equation (1) yields

percent reduction in apparent contrast

$$= 100\left[\frac{c_m - c_t}{c_m}\right] = 100\left[\frac{c_m - \dfrac{c_m}{1+b'}}{c_m}\right]$$

$$= 100\left[1 - \frac{1}{1+b'}\right]. \qquad (3)$$

One obvious implication of equation (3) is that the percent reduction in apparent contrast should be independent of the contrast level c_m of the matching stimulus. This can be checked against the available data: $c_s = 1$, $c_m = 0.5$ (Fig. 5) and $c_s = 1$, $c_m = 0.1$ (Fig. 6). There is a tendency, quite large in some instances (e.g. subject, JS, 3.3 c/deg) for a smaller reduction in apparent contrast to be associated with higher levels of c_m. The observed variation of percent reduction in apparent contract with c_m requires an elaboration of the simple theory.

An approximation to a theory of fully reciprocal interactions. A quite natural elaboration of the theory of equation (2) is to consider that not only does the surround inhibit the center but the center reciprocally inhibits the surround. Because of its smaller size and contrast, the center may exert less effect on the surround than vice versa. A first-order approximation to this reciprocal theory is simply to elaborate the term b' [equation (2)] to a b (no prime) that incorporates reciprocal inhibition from the center:

percent reduction in apparent contrast

$$= 100\left[1 - \frac{1}{1+b}\right]$$

$$= 100\left[1 - \frac{1}{1 + \dfrac{\sum_i w_{i,\theta} g_\theta(c_s)}{1 + h(c_m)}}\right] \qquad (4)$$

and to use this b instead of only its numerator [b' in equation (3)]. The function $h(c_t)$ is a monotonic increasing function that represents the inhibitory effectiveness of the center as a function of its output magnitude.

Equation (4) is an approximation because it uses only the first two terms of an infinite series of indirect effects in which the reciprocal feedback of the center affects the surround which affects the center, etc. Indeed, the situation is far more complex. The center is represented by a large aggregate of diverse neurons, as is the surround. Every neuron is involved with all of its neighbors in reciprocal feedback interactions. This fully interactive model is far beyond the scope of the present paper, both in complexity and in the number of assumptions that would be needed to fully specify the model. So, we stop with the first two terms. In this two-term approximation, the effects of varying the contrast of the center (which are represented in the denominator) are separable from the effects of varying the contrast of the surround (which are represented in the numerator). The function h absorbs the effect of *level* of matching contrast c_m on percent reduction in apparent contrast.

Orientation specificity. The surround, of course, has the biggest role in determining the percent reduction of apparent contrast of the center. We now consider the complex effects of surround contrast c_s, spatial frequency f of the center and surround, and relative orientation (||, ⊥) of center and surround. These are mediated by the functions $g_\parallel(c_s)$ and $g_\perp(c_s)$. The data allow us to distinguish between three complementary explanations of the relationship between inhibitory connections between pairs of texture-sensitive units with parallel receptive fields and pairs of texture-sensitive units with perpendicular receptive fields (see Fig. 8).

(i) *Early saturation*: $g_\parallel(c_s) = g_\perp(c_s)$ and $w_\perp = k_1 w$, $0 < k_1 < 1$ [Fig. 8(a)]. The function $g_\parallel(c_s)$ mapping input contrast to lateral inhibition for parallel surrounds and the function $g_\perp(c_s)$ for connections between units tuned to perpendicular surrounds are identically the same, only their weights differ. The functions saturate at contrasts $< \pm 1$.

(ii) *Low efficiency (same intercept)*: $g_\parallel(c_s) = g_\perp(k_2 c_s)$ and $w_\perp = w_\parallel$, $0 < k_2 < 1$ [Fig. 8(b)]. The function mapping contrast to lateral inhibition reaches the same maximum level for connections between units tuned to different orientations as it does for connections between units tuned to equal orientations, but it has a smaller slope (lower efficiency) for connections between units tuned to different orientations than it does for connections between units tuned to equal orientations.

(iii) *Low efficiency (non-saturating)* [Fig. 8(c)]. The linear functions shown in Fig. 8(c) satisfy the conditions on g and w defined in both (i) and (ii). That is, the function mapping contrast to lateral inhibition is *strictly* increasing. It reaches a different maximum level for connections between units tuned to different orientations than it does for connections between units tuned to equal orientations, and it has a smaller slope (lower efficiency) for connections between units tuned to different orientations than it does for connections between units tuned to equal orientations.

While each of these assumptions about the nature of $g_\parallel$ and $g_\perp$ can account for much of the data, none of them accounts for all. We consider now empirical criteria which, when are satisfied, would refute each of these interpretations. One way to refute early saturation is to demonstrate that, at high levels of surround contrast (e.g. $c_s = 1.0$) there is no indication of orientation specificity. To refute low efficiency (same intercept), it is sufficient to demonstrate that, at high levels of surround contrast there is distinct orientation specificity.

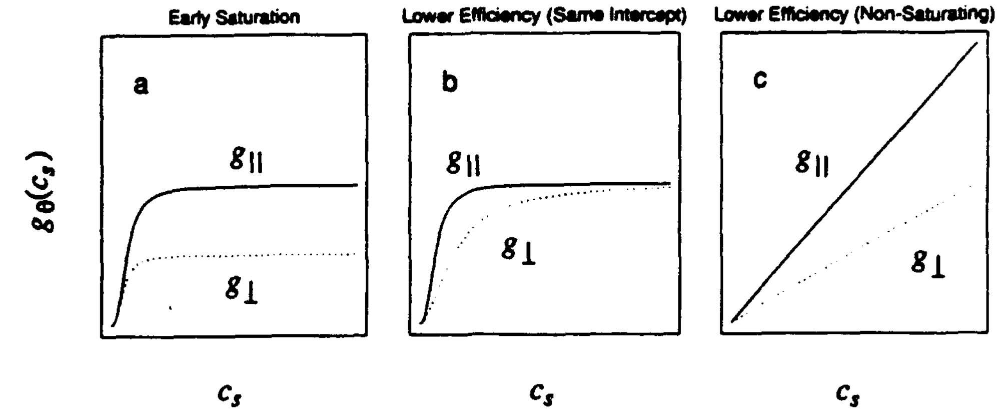


FIGURE 8. Three complementary relationships between the inhibition delivered by lateral connections to neurons of equal and different orientations: (a) Early saturation, the function mapping contrast to lateral inhibition has a lower intercept (earlier saturation) for connections between neurons tuned to different orientations (dashed line) than for connections between neurons tuned to equal orientations (solid line). (b) Low efficiency (same intercept), the function mapping contrast to lateral inhibition reaches the same maximum level for connections between neurons tuned to different orientations as it does for connections between neurons tuned to equal orientations, but it has a smaller slope (lower efficiency) for connections between neurons tuned to different orientations (dashed line) than it does for connections between neurons tuned to equal orientations (solid line). (c) Low efficiency (non-saturating), the function mapping contrast to lateral inhibition is *strictly* increasing, reaches a different maximum level for connections between neurons tuned to different orientations as it does for connections between neurons tuned to equal orientations, and has a smaller slope (lower efficiency) for connections between neurons tuned to different orientations (dashed line) than it does for connections between neurons tuned to equal orientations (solid line).

Low efficiency (non-saturating), can be refuted by demonstrating that, for a given c_m, an increase in surround contrast does not result in any increase in percent reduction in apparent contrast.

For 20.0 c/deg stimuli, only one value of c_s was tested, so we cannot refute low efficiency (non-saturating). However, we can refute low efficiency (same intercept) because, for both subjects, distinct orientation specificity is apparent in the data (Fig. 5).

For 10.0 c/deg stimuli again we are able to refute low efficiency (same intercept). There is distinct orientation specificity when $c_s = 1.0$ for both subjects, especially when $c_m = 0.1$ (Fig. 6). For 10.0 c/deg stimuli we are also able to refute low efficiency (non-saturating). There is no appreciable difference between the data from the $c_s = 1.0$, $c_m = 0.1$ viewing condition and the $c_s = 0.2$, $c_m = 0.1$ viewing condition, for either subject.

For 3.3 c/deg stimuli, however, things are much less clear cut. Both subjects' data display distinct increases in percent reduction in apparent contrast with an increase in c_s. This prohibits us from discrediting low efficiency (non-saturating). For JS, only with $c_m = 0.03$ does there appear to be some orientation specificity, when surround contrast is maximal. Whether or not this orientation specificity is distinct enough to refute low efficiency (same intercept) is a matter for debate. The most parsimonious judgment is to accept all three explanations as possibilities. For AS, only with $c_m = 0.1$ does there appear to be any significant amount of orientation specificity, when $c_s = 1.0$. Here again the best policy is not to discredit any of the three explanations. A summary of the possible explanations for each subject's data, at each spatial frequency, is given in Table 2.

TABLE 2. Possible explanations of orientation specific lateral inhibition

	Spatial frequency (c/deg)		
Subject	3.3	10.0	20.0
AS	ES; LE(SI); LE(NS)	ES	ES; LE(NS)
JS	ES; LE(SI); LE(NS)	ES	ES; LE(NS)

Explanations not discredited by the data are given in each cell.
ES, early saturation; LE(SI), low efficiency (same intercept); LE(NS), low efficiency (non-saturating).

CONCLUSION

Chubb *et al.* (1989) demonstrated that the lateral inhibition of perceived textural contrast is mediated by arrays of neurons that are narrowly tuned for spatial frequency. The results of these experiments indicate that they are tuned for orientation as well. This research also clearly indicates that the mechanism responsible for the lateral inhibition of perceived textural contrast receives equal inputs from both the on-center and the off-center visual pathways.

REFERENCES

Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.

Carlson, C. R., Anderson, C. H. & Moeller, J. R. (1980). Visual illusions without low spatial frequencies. *Investigative Ophthalmology and Visual Science, 19*, 165–166.

Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, 5*, 1986–2006.

Chubb, C. & Sperling, G. (1989). Second-order motion perception: Space-time separable mechanism. In *Proceedings of the Workshop* *on Visual Motion, 20–22 March 1989, Irvine, Calif.* (pp. 126–138). Washington, D.C.: IEEE Computer Society Press.

Chubb, C., Sperling, G. & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Science U.S.A.*, *86*, 9631–9635.

Enroth-Cugell, C. & Jakiela, H. G. (1980). Suppression of cat retinal ganglion cells by moving patterns. *Journal of Physiology*, *302*, 49–72.

Enroth-Cugell, C. & Robson, J. G. (1984). Functional characteristics of cat retinal ganglion cells. *Investigative Ophthalmology and Visual Science*, *25*, 250–267.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181–197.

Landy, M. S., Cohen, Y. & Sperling, G. (1984). HIPS: Image processing under UNIX. Software and applications. *Behavior Research Methods, Instruments and Computers*, *2*, 199–216.

Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early visual mechanisms. *Journal of the Optical Society of America A*, *7*, 923–932.

Ohzawa, I., Sclar, G. & Freeman, R. D. (1985). Contrast gain control in the cat's visual system. *Journal of Neurophysiology*, *54*, 651–667.

Sagi, D. & Hochstein, S. (1985). Lateral inhibition between spatially adjacent spatial frequency channels? *Perception & Psychophysics*, *37*, 315–322.

Schiller, P. H., Sandell, J. H. & Maunsell, J. H. R. (1986). Functions of the on and off channels of the visual system. *Nature*, *322*, 824–825.

Shapley, R. M. & Victor, J. D. (1978). The effect of contrast on the transfer properties of cat retinal ganglion cells. *Journal of Physiology*, *285*, 275–298.

Solomon, J. A. & Sperling, G. (1993). Full-wave and half-wave rectification in 2nd-order motion perception. *Vision Research*. In press.

Sperling, G. (1970). Model of visual adaptation and contrast detection. *Perception & Psychophysics*, *8*, 143–157.

Sperling, G. (1989). Three stages and two systems of visual processing. *Spatial Vision*, *4*, 183–207.

Sperling, G. & Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America*, *58*, 1133–1145.

Zemon, V., Gordon, J. & Welch, J. (1988). Asymmetries in on and off visual pathways of humans revealed using contrast-evoked cortical potentials. *Visual Neuroscience*, *1*, 145–150.

Acknowledgement—The research was supported by the Air Force Office of Scientific Research, Visual Information Processing Program, grant AFOSR 91-0178.

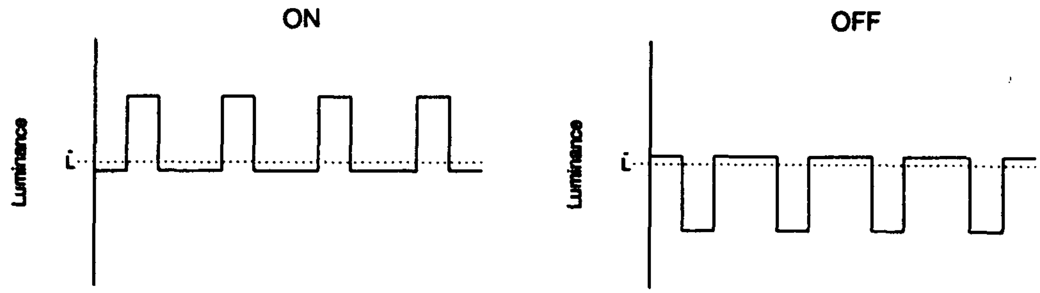


FIGURE 2. Illustration of "ON" and "OFF" textures. A vertical (or horizontal) slice through each texture is diagrammed. Mean luminance $\bar{L}$, is indicated on the ordinates.

Insofar as the on-center and off-center ganglion cells can be modeled as having center–surround antagonism and a (soft) threshold for firing, then the bright spots in our textures will selectively increase the firing rates of on-center cells in whose receptive field centers they fall, and the dark spots will increase the firing rates of off-center cells. Various plausible assumptions about the responsiveness of on- and off-center systems make a high

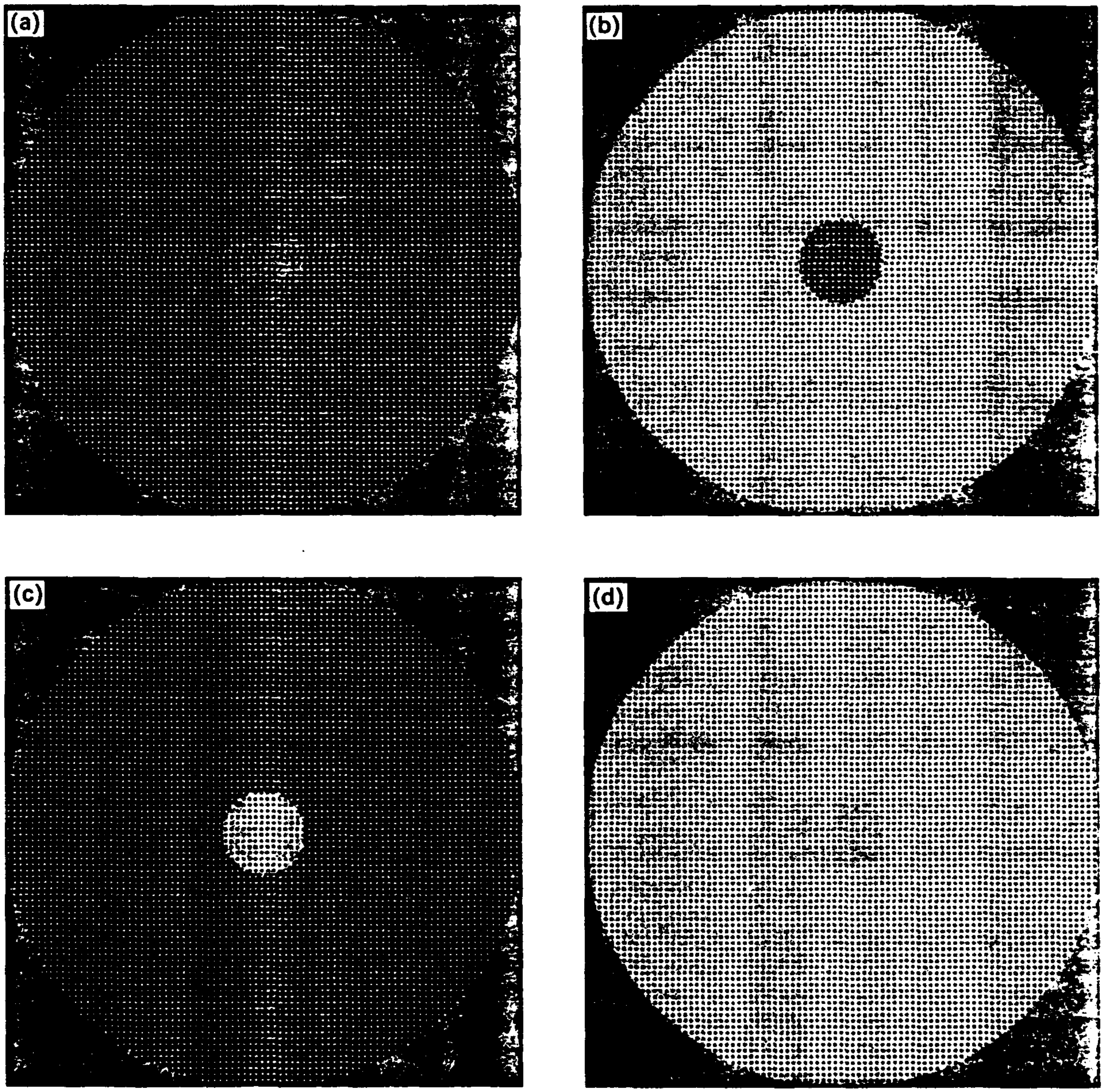


FIGURE 3. Stimuli for Expt 1. (a) On-center stimulating center, on-center stimulating surround (ON/ON). (b) On-center stimulating center, *off-center stimulating surround* (ON/OFF). (c) Off-center stimulating center, on-center stimulating surround (OFF/ON). (d) Off-center stimulating center, off-center stimulating surround (OFF/OFF).

Journal of Experimental Psychology:
Human Perception and Performance
1993, Vol. 19, No. 4, 845–866

0096-1523/93/$3.00

Information Transfer in Iconic Memory Experiments

Karl R. Gegenfurtner and George Sperling

To report letters from briefly exposed letter arrays, subjects must transfer information from a rapidly decaying trace (iconic memory) to more durable storage. In a partial-report paradigm, we systematically varied the proportion (P) of trials with a long cue delay relative to a short cue delay. Practiced subjects used the same transfer strategy independent of P. Data from a partial-report-plus-masking experiment were used to construct a computational model that accurately predicted partial- and whole-report performance with and without masks. Assumptions: Prior to a cue, subjects attend primarily to the middle row of a three-row display, resulting in nonselective transfer. After the cue, they attend only to the cued row. Transfer rate is the product of iconic legibility (which depends on time and retinal location) and attention allocation (which shifts after a cue). Cumulative transfer is limited by the capacity of durable storage.

When subjects are asked to report all the letters they can see in a brief flash of a letter array, they usually can report only four or five letters. The number of reported letters is independent of the number of displayed letters (when more than about five letters are displayed; e.g., Sperling, 1960). One might therefore infer that the limit on the number of letters reported is due to a limited memory capacity, traditionally called the "span of apprehension" (Külpe, 1904; Wundt, 1899). However, a partial-report procedure demonstrates that subjects are able to store a dozen or more items in a very short-term memory (Sperling, 1960).

In a typical partial-report experiment, a 3 × 3 letter matrix is followed by a cue (e.g., a high-, middle-, or low-pitched tone) that indicates the row of the matrix that the subject has to report. Figure 1 shows results from a partial-report experiment. When the cue occurs at the same time as the letters or shortly afterwards, the subject can report all the letters in the cued row. Because the subject does not know in advance which row will be cued, perfect performance implies that all the items are stored and still available at the time of the cue. When the cue is delayed, partial-report performance decreases, until it finally reaches the level of whole report at cue delays of about 500–800 ms.

The decay of partial-report accuracy with cue delay has been taken as evidence for a second kind of memory, which Neisser (1967) called "iconic memory." Neisser assumed that initially all items are held in iconic memory and, at the time of the cue, the cued letters are transferred into a longer lasting storage.

Durable Storage

The partial-report experiment itself does not prove the existence of two different memories. The cue delay effect might be caused by one type of memory, which decays to four or five letters, and the subject has control over which letters survive. However, experiments with a poststimulus mask (Averbach & Sperling, 1960; Sperling, 1960, 1963) show that there is more than one memory. When a poststimulus mask comes soon after stimulus offset, there is a marked decrease in performance relative to the no-mask control condition. Therefore the storage that is probed in the partial-report experiment is destroyed by the mask. When a poststimulus mask comes later, say, 1 s after the stimulus, it does not influence partial reports at all. The interaction of mask and cue delay implies that there are at least two types of memory. One, iconic storage, has a large capacity, decays rapidly, and is destroyed by a mask following the stimulus. The other storage can hold only a limited number of items but is not affected by masking and seems to have a long lifetime. Following Coltheart (1980), we call the second type of memory "durable storage."

There have been several attempts to discriminate among types of durable storage in partial-report and similar types of experiments (Adelson & Jonides, 1980; DiLollo, 1984; Duncan, 1983; Irwin & Yeomans, 1986; Kaufman, 1978; Loftus, Johnson, & Shimamura, 1985; Mewhort, Campbell, Marchetti, & Campbell, 1981; Scarborough, 1972; Sperling, 1967; Townsend, 1973). Similarly, there have been attempts to discriminate between iconic memory as revealed by partial-report experiments, which use informational measures, and other kinds of visual sensory memory that might be revealed by direct sensory judgments, which use synchrony judgments of visual traces with auditory clicks

Karl R. Gegenfurtner, Howard Hughes Medical Institute and Center for Neural Science, New York University; George Sperling, Psychology Department and Center for Neural Science, New York University.

The experimental work was supported by the U.S. Air Force Office of Scientific Research, Life Science Directorate, Visual Information Processing Program; the theoretical work and preparation of this article were supported by Office of Naval Research, Cognitive and Neural Sciences Division.

A preliminary report of Experiments 1 and 2 was made by Sperling and Gegenfurtner (1988).

Correspondence concerning this article should be addressed to George Sperling, who is now at the Department of Cognitive Science, University of California, Irvine, California 92717. Electronic mail may be sent to sperling@uci.edu.

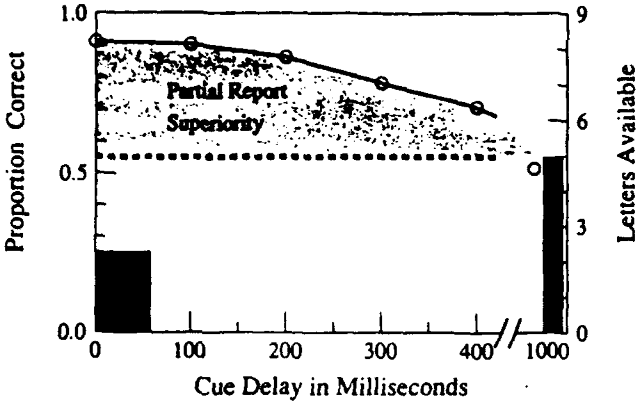


Figure 1. Method and results of a typical partial-report experiment. (The abscissa is the time, relative to the onset of the letter array, of a tonal cue to report a given row. The ordinate is the mean proportion of correctly reported letters in the cued rows. Open circles represent the performance of Subject BL. The filled bar on the right indicates performance in a whole-report experiment. The shaded area between the dashed line and the dots represents partial-report superiority [relative to whole report]. The rightmost data point indicates performance measured at a cue delay of 1,000 ms, which is comparable to the delay involved in the whole-report procedure.)

(Efron, 1970; Sperling, 1967), the integration time between successive stroboscopic flashes (Eriksen & Collins, 1967; Hogden & DiLollo, 1974), and many other procedures (see reviews by Coltheart, 1980; Long, 1980). Because we are concerned only with partial-report experiments here, we can bypass these issues and deal only with iconic memory and durable storage without any further specifications or subdivisions.

Selective and Nonselective Transfer

In a typical iconic memory experiment, at intermediate cue delays, the quality of the iconic image has deteriorated so that few if any additional items can be transferred into durable storage. If the cue were to be further delayed until the iconic image had decayed completely, no transfer from the cued row would be possible, and performance would decrease to 0. As shown in Figure 1, this is not the case. Performance at long cue delays reaches asymptote at whole-report performance, not at zero.

The failure of partial-report accuracy to decay to zero as a function of cue delay is almost certainly due to what Averbach and Coriell (1961) called "nonselective readout." In the time between stimulus and cue, subjects start to transfer items from iconic memory to durable storage. This transfer is nonselective with respect to the cue. It saves subjects from performing badly at long cue delays. On the other hand, the letters transferred nonselectively use up some of the limited capacity of durable storage. By the time the cue appears, durable storage might already be filled with items from the noncued rows. For example, about four items can be transferred in the first 100 ms after stimulus display (Sperling, 1963, 1967). This is approximately the capacity of durable storage. Thus the strategy of nonselective readout at short cue delays might have the disadvantage of overcrowding durable storage, thereby slowing down the subsequent transfer from the cued row.

As several authors (Coltheart, 1980; Dick, 1969; Hall, 1974; Merikle, Lowe, & Coltheart, 1971; Mewhort, Johns, & Coble, 1991; Mewhort, Merikle, & Bryden, 1969; Sakitt, 1976; Sperling, 1960; Sperling & Dosher, 1986) have suggested, it is possible that subjects deliberately use different strategies at different cue delays. On trials with long cue delays, subjects use nonselective readout, to avoid the disaster of having iconic memory decay to illegibility before any items have been transferred. At short cue delays, subjects pay equal attention to all rows and do not transfer items nonselectively, to avoid filling durable storage. Of course, selective strategies would be possible only when partial-report experiments are run in a blocked design, as they typically have been (e.g., Irwin & Yeomans, 1986; Mewhort et al., 1981; Sakitt, 1976; Sperling, 1960). In each block of trials, only one cue delay was used; thereby, subjects can use the strategy that is most advantageous for the particular cue delay.

Overall Plan

In our first experiment, we attempted to discriminate between short-cue-delay and long-cue-delay coding strategies that subjects might use in iconic memory experiments and to determine the costs and benefits of each strategy. It was essential to resolve the possibility that subjects tailor their strategy to the particular cue delay before we proceeded with Experiment 2, a parametric investigation of partial-report accuracy in 25 combinations of Cue Delay × Mask Delay. The data of Experiment 2 enable us to define a model that mathematically describes the time courses of iconic decay and the twin processes of selective and nonselective retrieval. The model, which aggregates all the rows of a three-row stimulus, is very successful computationally, but it contains two enigmas. These enigmas are resolved by noting that characters from the middle row are transferred to durable storage much more rapidly than characters from the other rows and embodying this fact in an elaborated model that treats each row separately. We consider how our model differs from prior computational approaches to iconic transfer. Finally, we show that our formulation of the role of attention in selective transfer is consistent with many other attentional phenomena.

Experiment 1: Coding Strategies

Suppose that subjects in an iconic memory experiment use a strategy of selective transfer at short cue delays and a strategy of nonselective transfer at long cue delays. We wished to estimate the cost of each strategy when it was used inappropriately and the benefit of each strategy when it was used appropriately in terms of the cost–benefit analysis of Posner, Nissen, and Ogden (1978). The problem in estimating the cost of nonselective transfer at short cue delays was to get the subject to use nonselective transfer (an inappropriate strategy) at short cue delays. Here is the trick. In a blocked situation with only long cue delays, we assumed subjects would certainly use nonselective transfer (the appropriate long-delay strategy). When we included very few trials with a short cue delay in such a block, subjects were still better off from a decision-theoretic viewpoint using nonselective transfer throughout the whole session. Assuming that subjects would try to maximize their performance, we predicted they would use the long-delay strategy in blocks of predominantly long delays and the short-delay strategy in blocks of predominantly short delays. The difference in performance between (a) the few short-delay trials embedded in a block of predominantly long-delay trials and (b) a pure block of short-delay trials provided an estimate of the cost of using nonselective transfer (the inappropriate, long-delay strategy) at short cue delays.

A similar argument was posited for long cue delays. We predicted that when a short cue delay was presented 95% of the time, subjects would use selective transfer (the appropriate strategy). On the occasional long cue delays, we expected their performance to be very poor. Figure 2 illustrates a predicted outcome of this sort of experiment. Performance at long cue delays was expected to decrease markedly with a reduction in the probability of the occurrence of a long cue delay. Performance for both types of cue delay was expected to be highest in the blocked design condition.

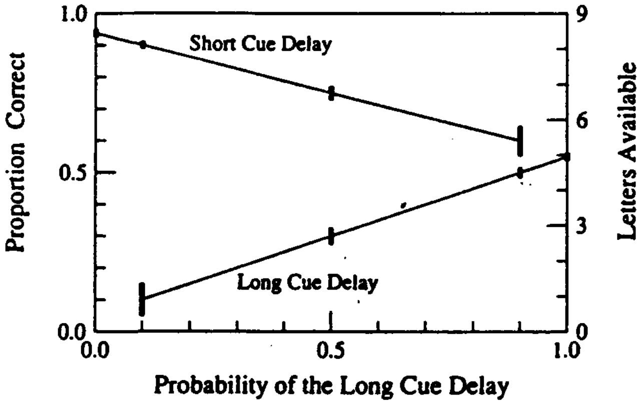


Figure 2. Strategy analysis: The expected outcomes of the cost analysis experiment in which either a long or a short cue delay can occur on a trial. (The abscissa is the probability, within a block of trials, of the long cue delay. The ordinate is the mean proportion of correct reports, conditioned on the type of cue [long vs. short] delay. The upper curve is the expected performance with short cue delays; the lower curve represents long cue delays. Performance at long cue delays is expected to be poor when long cue delays occur rarely [bottom left] and to increase as they become predominant. Performance at short cue delays is symmetrically opposite. The solid bars through the data points indicate standard errors. Their differing length indicates that in the low-probability conditions, fewer trials will be available.)

Method

Subjects. Two graduate and two undergraduate students at New York University participated in the experiment for pay. All subjects had normal or corrected-to-normal vision. Each subject had a minimum of five practice sessions of 200 trials each; for some subjects, practice continued longer until their performance in a regular partial-report experiment with equally likely cue delays reached a steady state. Subjects BL and PC were presented 3 × 3 arrays. Performance for subjects RS and BF was better, so they were shown 3 × 4 arrays.

Stimuli. All experiments were controlled by a Digital Equipment Corporation PDP-11/23 computer. The letters were presented on a Hewlett Packard 1310A cathode ray tube (CRT) with a fast white P4 phosphor. The CRT was driven by a specially designed display interface (Kropfl, 1975) and software for real-time vision experiments (Melchner & Sperling, 1980). Tones were presented on Sennheiser HD414 headphones. A Wavetek Model 159 waveform generator was used to generate the tones, which were set to a comfortable listening level. The timing of the actual stimulus sequences was verified by independent oscilloscopic measurements and was accurate to within 1 ms.

The stimuli consisted of a 3 × 3 or 3 × 4 array of letters. Figure 3a shows a photograph of a typical stimulus. The whole display extended 3.1° or 4.5° of visual angle, respectively, at a viewing distance of 128 cm. Each letter was 1.2 cm high and 1.0 cm wide, with a distance between letters of 2.0 cm horizontally and 1.8 cm vertically. Viewing was binocular.

The luminance of the letters was determined by measuring the luminance of a uniform rectangle with a United Detector Technologies photometer, which had been calibrated against a standard light source. The rectangle had the same pixel intensity as the letters, the same pixel spacing, and the same number of dots as the letter bitmaps. The measured luminance was 34 cd/m^2. The letters were displayed on a dark background of approximately 0.05 cd/m^2. The room was dimly illuminated, and the wall behind the monitor had a luminance of approximately 1.2 cd/m^2. The individual letters were randomly chosen without replacement from the set of 20 consonants, excluding *Y*.

Procedure. Each partial-report session consisted of 200 trials. Figure 4a shows a flow diagram for one trial. The subject initiated the trial by pressing a button. After a random interval of 1.0–1.5 s, the stimuli were displayed for 50 ms (five repeated frames at 10 ms per frame). At the time specified by the cue delay, a tone was sounded on the headphone for 100 ms. The frequencies of the cue tones were 225, 600, and 975 Hz for the bottom, middle, and top row, respectively. The time for the cue delay was measured from the onset of the stimulus. Typically, cue delays in partial-report experiments have been specified in terms of the time from stimulus termination (e.g., Sperling, 1960, and many others). Our reason for specifying a cue delay relative to stimulus onset was that the delay then corresponded to the time for which stimulus information was available before a cue appeared.

Cue delay could be varied independently of the other stimulus parameters, and the cue could occur before stimulus onset, during the stimulus, or after stimulus termination. After the stimulus sequence, the subject was prompted on the screen for a typed response. After the subject responded, the correct letters were shown on the screen, together with the subject's response. Then the next trial started. A response letter was scored as correct only when it was reported in the correct serial position.

In this experimental design, it is inevitable that the low-probability condition for one cue delay coincides with a high probability for the other cue delay. Therefore the number of observations

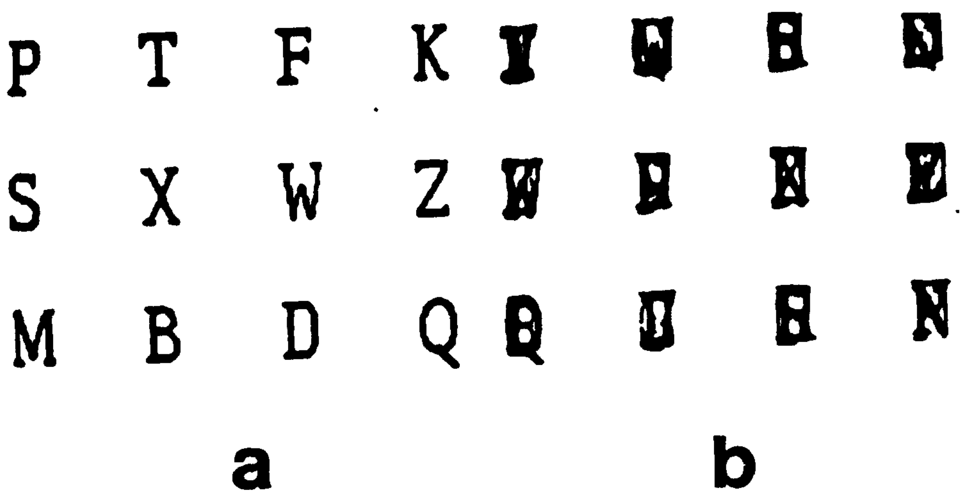


Figure 3. Panel a shows a typical stimulus display. (In both Experiments 1 and 2, all letters were white on a black background; their contrast is reversed here for better reproducibility. Subjects BF and RS used a 3 × 4 matrix of letters; Subjects PC and BL used a 3 × 3 matrix.) Panel b shows a mask. (Like the stimulus, the mask is shown in reversed contrast. Each component masking pattern consists of five different letters displayed in extremely rapid succession.)

for the low-probability conditions is smaller. Performance was evaluated for cue delay probabilities of 0.1, 0.5, 0.9, and 1.0. The short cue delay used was always 0 ms. The duration of the long cue delay was chosen separately for each subject so as to achieve a performance level that would still be better than whole report. The delay values were 400, 800, and 1,000 ms.

Results

The data were analyzed separately for each subject. Figure 5 illustrates the results for subjects PC and BF. Table 1 summarizes the data for all 4 subjects. If there is no effect of probability of occurrence, then performance should not vary and all data points for a fixed cue delay should fall on a straight horizontal line. We therefore estimated slope and intercept of the best fitting lines (in the least squares sense) through the data.

Table 1 shows that the slopes of least-squares-estimated lines through the data points were all negligibly small. Seven of eight slopes were negative, and none of them was significantly different from 0 (according to *t* tests at the .05 significance level).

Discussion

Three unambiguous aspects of the data lead to three significant conclusions.

Performance in response to a low-probability long-delay cue did not approach zero but reached asymptote at a level typical for whole report. This means that subjects always used nonselective transfer.

Performance in response to a low-probability short-delay cue was not impaired compared with that in response to a high-probability short-delay cue. We infer that nonselective transfer did not involve any additional cost for the subject, even on trials in which selective transfer was also used.

The finding that performance was better for short- than long-delay cues indicates that subjects indeed used selective transfer for short-delay cues.

Finally, with respect to experimental procedures, if subjects have more than one good strategy available, the particular mixture of strategies that they use would depends on the particular mixture of cue delays they confront. The finding that subjects used the same strategy for short- and at long-delay cues greatly simplified the design of Experiment 2. The mixture of cue delays could be optimized for obtaining the desired data, unconstrained by (non)effects on subjects' strategies.

The finding that a single transfer strategy was used at all cue delays is in striking contrast to previous observations suggesting at least two strategies. Sperling (1960, Figure 5) showed a subject whose short-delay-cue strategy failed against long-delay cues and whose long-delay-cue strategy failed to take advantage of short-delay cues. Another (famous) subject (Sperling, 1990, Figure 6) retained his short-delay-cue strategy for too long a cue delay, thereby producing a nonmonotonic iconic decay function. The simplest explanation for the discrepancy between the present data and Sperling's data is that the earlier data were obtained in the first few hundred trials with naive subjects whose performance was clearly nonoptimal. The present data show that, after practice, subjects acquire a single strategy that is effective for both long and short cue delays.

Experiment 2: Time Course of Iconic Memory

The results of Experiment 1 left us with two open questions: How do subjects avoid overfilling durable storage when selective transfer follows nonselective transfer? More generally, how are nonselective transfer and selective transfer combined? To address these questions, we introduced a

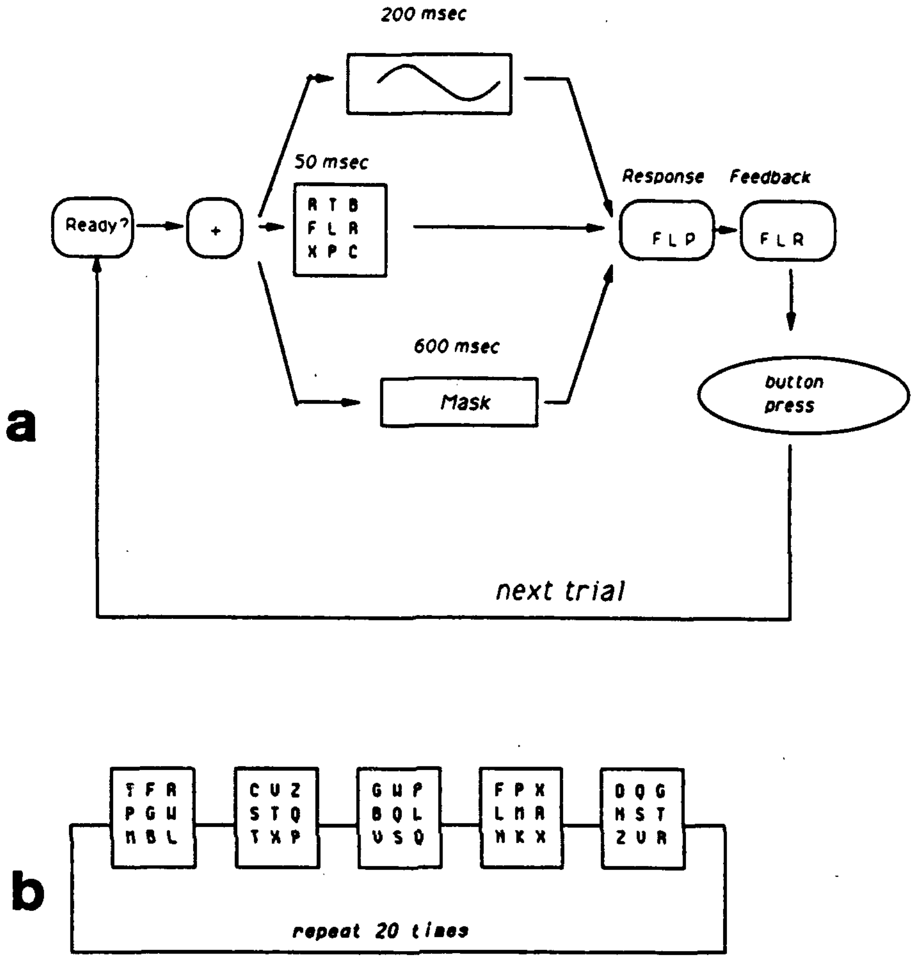


Figure 4. Panel a is a flow chart for a trial. (The three parallel streams for letters, cue, and mask indicate that the onset times for these could be varied independently to produce any arbitrary ordering. The mask was used in Experiment 2 only.) Panel b is a flow chart for the production ot a mask. (A sequence of five different frames is painted with 6-ms interframe intervals; the sequence of five is repeated 20 times.)

variably delayed poststimulus mask into the partial-report procedure.

An appropriately chosen visual postexposure masking stimulus should have two properties: It should destroy the contents of iconic memory but leave durable storage unimpaired. For the destruction of iconic memory, a mask is constructed in such a way that when it and the test stimulus are exposed simultaneously, the test stimulus is masked to the point of unintelligibility (Kahneman, 1968; Sperling, 1963). The ability to mask the test stimulus completely when it is strongest (i.e., when it is physically present) implies that the postexposure masking stimulus will even more effectively mask the test after it has been weakened by decay. The ability to leave durable storage unimpaired is demonstrated by showing that long mask delays yield equivalent performance to no-mask control conditions. Such a poststimulus mask serves to limit the time for which information from iconic memory is available for transfer to durable storage. By varying cue delay and mask delay independently in a crossed design, we obtained estimates for the amount of transfer to durable storage in each interval.

Figure 6 illustrates the logic of the masking paradigm in three kinds of conditions. In the first condition (Figure 6a), the cue occurs after stimulus onset and before mask onset. During the interval between stimulus onset and the cue, the subject does not know which row will be cued. Therefore, all transfer is nonselective with respect to the cue. After the cue has occurred, the subject switches attention to the cued row and transfers letters selectively from that row. We call these two kinds of information transfer from iconic memory to durable storage nonselective and selective transfer, respectively.

Two special cases lead to pure selective and pure nonselective transfer. When the mask comes before or at the same time as the cue, only nonselective transfer occurs (Figure 6b). When the cue comes at or before stimulus onset, subjects use selective transfer throughout (Figure 6c). In all other cases (Figure 6a), there is a mixture of selective and nonselective

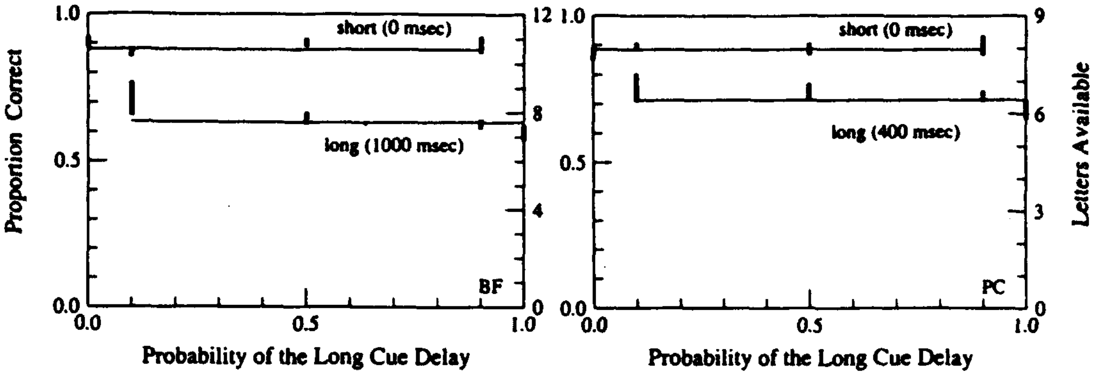


Figure 5. The absence of strategy effects. (Results of Experiment 1 are shown for Subjects PC and BF. The ordinate is the proportion of correct reports; the abscissa is the probability of the long cue delay in a block of 200 trials. The two lines in each panel are the best fitting horizontal lines to the data for each cue delay. The vertical bars represent the standard error [±1 σ].)

transfer. Because the cue is irrelevant to nonselective transfer, the pure nonselective conditions should yield the same results as a whole-report experiment with similarly delayed poststimulus masks. This whole-report experiment was carried out as a control condition.

Method

The general experimental methods and subjects were the same as in Experiment 1 except for the following changes.

Subjects. Two subjects, BL and BF, who had served in the Experiment 1, served again in Experiment 2. It should be remarked that BF was able to report one or two items more than average from brief visual exposures. This would place him in the upper 10–20% of subjects in our experience. He was persuaded to serve in this tedious experiment in our hope of discovering some other unusual ability. However, except for a slightly higher level of performance, his data were typical in all respects. In addition, the two other subjects from Experiment 1 served for about half as many trials as BL and BF. Their data did not differ in any important ways from those of Subjects BL and BF and are not presented here.

Masking stimulus. In Experiment 2, a masking pattern (the mask) was displayed at a specified mask delay, which could be shorter or longer than the cue delay. A mask consisted of five different letters displayed in extremely rapid succession at each spatial location, so that the letters were summed by the visual system and could not be recognized individually (Budiansky & Sperling, 1969). All the letters comprising one frame of a mask were painted within 6 ms, a new frame was presented every 6 ms, and the sequence of five different frames was repeated 20 times for a total mask duration of 600 ms. The flow diagram in Figure 4 illustrates this process. The intensity of masks, measured in the same way as the intensity of the letters in Experiment 1, was 47 cd/m^2. Figure 3b illustrates a typical masking pattern. In a brief control experiment, it was verified that recognition of a stimulus letter was at chance when it was presented at the same time as a mask.

Procedure. Mask delays of 100, 200, 300, 400, and 500 ms were used in the experiment. The cue delays chosen were 0, 100, 200, 300, and 400 ms. On each trial, a cue delay and mask delay were chosen randomly in a mixed-list design. Each subject was tested on approximately 5,000 trials in 45-min long sessions of 200 trials each.

Table 1
The Proportion of Correctly Reported Letters as a Function of the Probability of Cue Delays in Experiment 1

	Probability of cue delay								
Subject/cue delay (ms)	0.1	No. of observations	0.5	No. of observations	0.9	No. of observations	1.0	No. of observations	Slope
BF									
0	0.897	92	0.905	283	0.864	736	0.904	200	−0.015
1,000	0.719	64	0.648	317	0.63	708	0.598	200	−0.012
PC									
0	0.9	40	0.887	140	0.879	360	0.863	200	−0.026
400	0.757	40	0.743	140	0.727	360	0.663	200	−0.082
BL									
0	0.884	23	0.916	103	0.923	181	0.912	80	0.034
800	0.667	19	0.581	97	0.605	177	0.648	53	−0.026
RS									
0	0.983	35	—	—	0.956	244	0.992	32	−0.068
1,000	0.717	23	—	—	0.714	365	0.631	80	−0.06

Note. Slope data indicate the slope of a least squares fitted line through the data points.

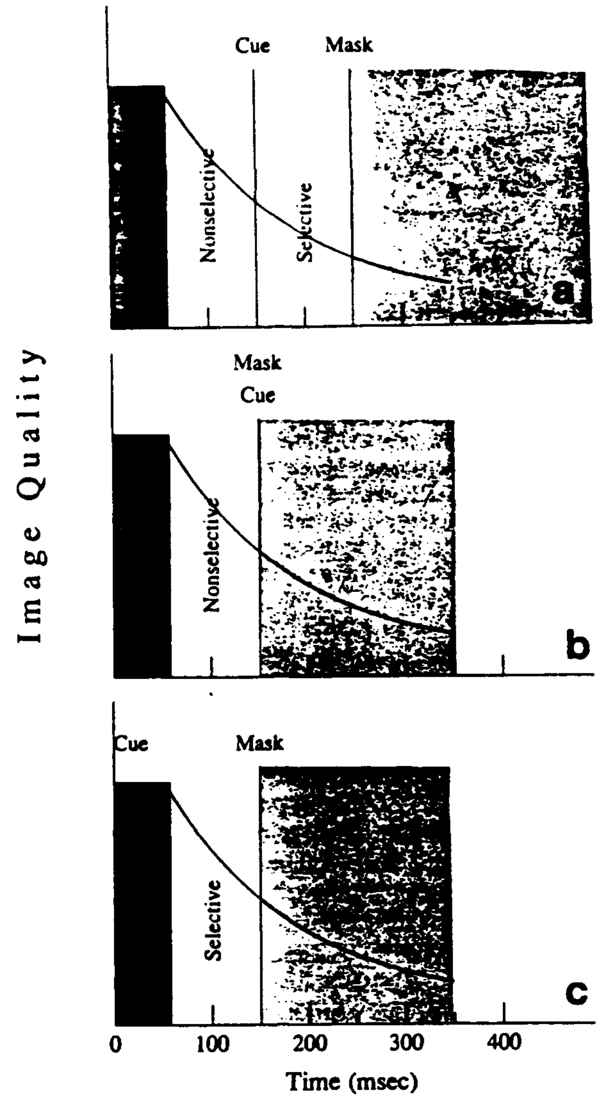


Figure 6. The logic behind the partial-report-plus-mask experiment. (Panel a shows nonselective and selective transfer. The cue occurs before the mask, nonselective transfer occurs before the cue, and selective transfer occurs during the interval cue to the mask. Panel b shows pure nonselective transfer. The cue occurs at or after the onset of the mask. Nonselective transfer ceases after onset of the mask; there is no resumption of transfer after the cue occurs. Panel c shows pure selective transfer: The cue comes at or before the onset of the stimulus.)

Whole report. In the whole-report condition, the subject was asked to report all the letters in the display. The same mask used in the partial-report condition was used. The whole-report practice and test conditions were run in separate sessions after the subjects were already practiced in the partial-report task. Data were collected only after performances had reached asymptote.

Results

As in Experiment 1, the data were analyzed separately for each subject. Figure 7 shows the effect of cue delay with mask delay as a parameter. As in other partial-report experiments, performance dropped as cue delay increased, confirming that subjects made efficient use of the cue. A strictly monotonic decrease in proportion correct as a function of cue delay means that using selective transfer in any time interval yielded more correctly reported letters than did nonselective transfer.

Figure 8 replots the data of Figure 7 with mask delay as the abscissa and cue delay as a parameter. The effect on performance of mask delay was also monotonic; the number of transferred letters increased rapidly with increasing mask delay. A monotonic increase in proportion correct with increasing mask delay means that additional available time for processing the stimulus was always useful.

Pure nonselective transfer. Figure 9 shows data for pure nonselective transfer—all the trials on which the cue occurred simultaneously with or after mask onset. Performance increased very quickly in the first 100 ms and then reached asymptote at around four or five letters. This indicates that these subjects were able to read about four letters in less than 100 ms, which is at the same level that other investigators have found (e.g., Sperling, 1963). Figure 9 also shows the data from the whole-report procedure. Whole-report accuracy is slightly lower than partial-report accuracy. We assume that this slight whole-report deficit was due to the larger number of letters that needed to be reported. Subjects might have occasionally forgotten a letter while reporting the earlier ones. Therefore the partial-report-plus-masking procedure seems to be a slightly better indicator of nonselective transfer than whole report.

The extreme right of Figure 9 shows that whole reports with a 500-ms mask yielded equivalent performance to that in the no-mask control condition. This result indicates that the masking stimulus satisfied the second condition stated for a successful mask: It did not interfere with the contents of durable storage.

Pure selective transfer. The subset of conditions with a cue delay of 0, which indicate pure selective transfer, yielded data that are superficially similar to nonselective-transfer data when graphed in terms of the actual number of letters reported, as shown in Figure 10. Accuracy increased monotonically with mask delay. However, selective transfer took longer than nonselective transfer to approach its asymptotic level (approximately 400 ms vs. 200 ms). The asymptotic accuracy level of selective transfer was much higher than that of nonselective transfer (90% vs. 50%), indicating a partial-report advantage. As in the case of nonselective transfer, when a mask was delayed 500 ms, there was only a negligible difference between mask and no-mask conditions.

An Aggregate-Row Model of Iconic Memory

Experiment 2 characterized purely selective and purely nonselective transfer. In an attempt to explain how they both combine in the overall transfer to durable storage, we developed a model that aggregates performance over rows. Subsequently we found that although the model gave excellent predictions of the present data, it left some serious residual problems. To resolve these, we developed a more

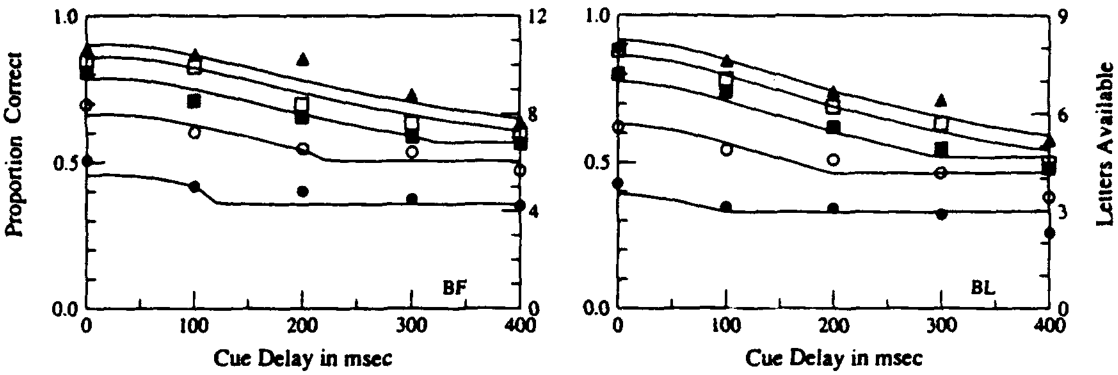


Figure 7. Accuracy of partial reports as a function of cue delay, with mask delay as the parameter: Experiment 2. (The ordinate indicates the proportion of correctly reported letters. The right ordinate indicates the corresponding number of letters transferred to durable storage. Each data point represents 150–250 trials. Panels indicate data for Subjects BL and BF: BL viewed 3 × 3 displays, and BF viewed 3 × 4 displays [Rows × Columns]. The curves drawn through the data points are the best fitting predictions of the two-process aggregate-row model described in the text.)

complicated model in which each row is considered separately. The formulation of the aggregate-row model is presented in this section.

Basic Assumptions: Additivity of Nonselective and Selective Transfer

Both selective and nonselective transfer contribute to the overall performance. In Experiment 2, only the contribution made by nonselective transfer was directly observable. The contribution of selective transfer could be observed only in the absence of nonselective transfer, that is, when selective transfer started immediately at stimulus onset with a cue delay of 0. We now estimate selective transfer at nonzero cue delays. We proceed by making an assumption about the combination rule for selective and nonselective transfer. This assumption allows us to subtract nonselective transfer from overall performance to derive selective transfer at various cue delays.

The simplest combination rule is additivity of the two transfer processes. (Averbach & Coriell, 1961, made a different assumption, which is considered in the Discussion.) To implement additivity of transfer processes, we make the following assumptions. (a) Letters are transferred nonselectively from stimulus onset on until the cue comes. (b) Selective transfer begins at onset of the cue and ends at onset of the mask, when all further information transfer out of iconic memory stops. (c) The total number of letters transferred is the sum of both transfer processes. Specifically, given a cue at time c and a mask at time m, the total number of letters, $L_{c,m}$, transferred from the cued row is the sum of the number of nonselectively transferred letters from the cued row, $(\frac{1}{3})N_{c,m}$, and the number of selectively transferred letters from the cued row, $S_{c,m}$:

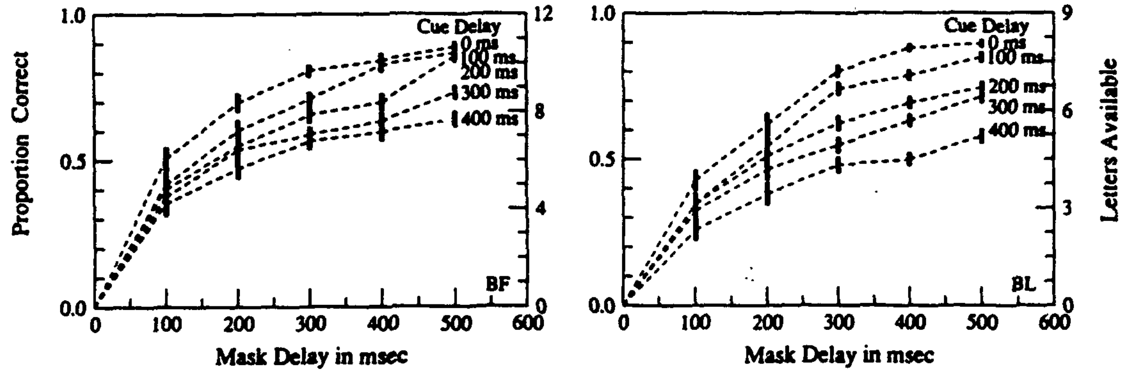


Figure 8. Results of Experiment 2 replotted to show the accuracy of partial reports as a function of mask delay, with cue delay as the parameter. (The vertical bars through the data points indicate the standard error of the proportions. The data points for each cue delay are connected by dotted lines. See Figure 7 for details.)

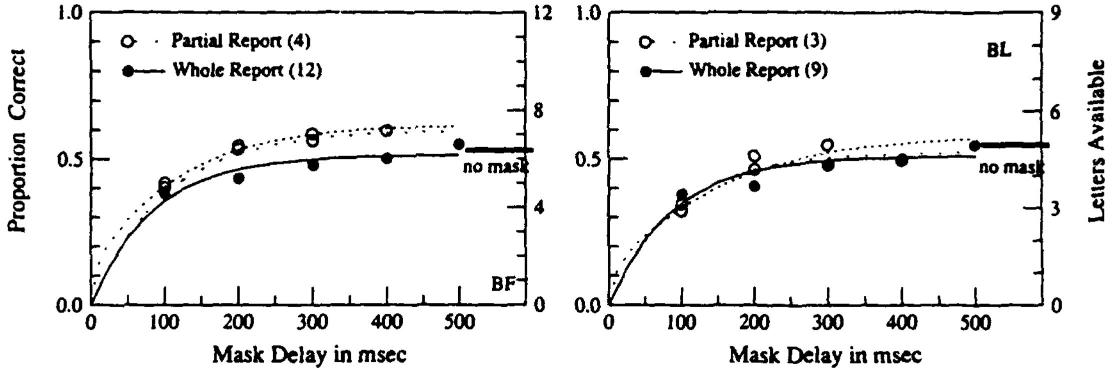


Figure 9. Pure nonselective transfer as a function of mask delay. (The open symbols are the proportion of correct partial reports on trials in which the cue occurred after mask onset [Figure 6b]; the filled symbols are the proportion of correct reports in whole-report-plus-mask trials. The horizontal bar at the right border indicates performance on whole-report trials without masks. The predictions of the aggregate-row model of iconic memory for partial and whole reports are indicated by solid and dotted lines, respectively. The dashed line shows the partial-report predictions of the model described in Equation 13, averaged over the three rows.)

$$L_{c,m} = S_{c,m} + \tfrac{1}{3}N_{c,m}. \qquad (1)$$

The factor ⅓ express the fact that because there are three equally likely cues, only one-third of the nonselectively transferred letters are expected to be in the cued row. To apply Equation 1 to our data, we note that we already know two of its three components. If we assume, for the moment, that all letters that are transferred from iconic memory to durable storage are reported, then the partial-report data directly yield the total reported letters, $L_{c,m}$. Partial reports made when the cue occurs simultaneously with or after the mask give the pure nonselective component, $\tfrac{1}{3}N_{c,m}$ $(c \geq m)$, the analysis illustrated in Figure 9. The difference between $L_{c,m}$ and $\tfrac{1}{3}N_{c,m}$ is the selective transfer, $S_{c,m}$.

Figure 11 shows the values of selective transfer derived from our data. Note that the only difference between Figures 8 and 11 is that the nonselective transfer component has been subtracted from overall performance. All the curves for selective transfer appear to be parallel, shifted vertically. This implies that only one factor determines selective transfer—time elapsed since stimulus onset. Selective transfer that begins, for example, 200 ms after stimulus onset will transfer just as many items in the time period from 200 to 500 ms as selective transfer that began at 0 ms. Because the rate of selective transfer depends only on the elapsed time since stimulus onset, it directly reflects the quality of the stimulus information.

To test the assumption of additivity, we fit the best set of perfectly parallel curves to our data. We do not make any assumptions about the form of the selective or nonselective transfer curves. The solid line segments in Figure 11 all derive from a single curve that has been translated up or down.

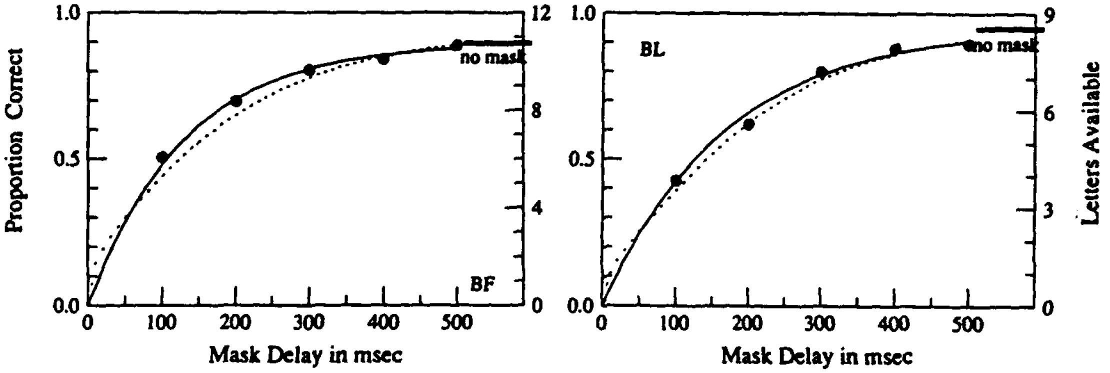


Figure 10. Pure selective transfer as a function of mask delay. (The filled symbols show the proportion of correct partial reports on trials on which the cue occurred at stimulus onset [Figure 6c]. The solid line shows the prediction of the aggregate-row model of iconic memory. The horizontal bar on the right border shows observed performance in a partial-report experiment without masks and a cue delay of 0. The solid line shows the predictions of the model described in Equation 14, averaged over the three rows.)

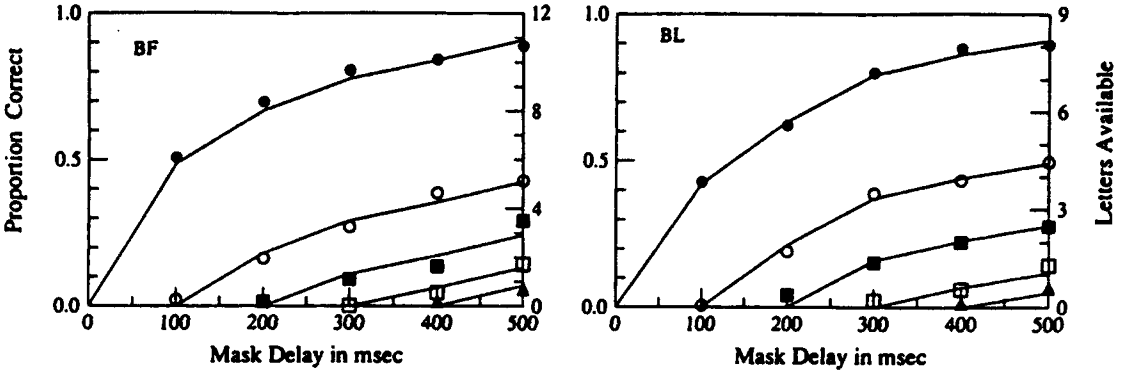


Figure 11. Test of the additivity assumption in the aggregate-row model. (The curves are based on Figure 8—accuracy of partial reports as a function of mask delay, with cue delay as the parameter. Cue delays are in milliseconds: Filled circle = 0, open circle = 100, filled square = 200, open square = 300, triangle = 400. The assumption of algebraic additivity of transfer [Equation 2] permits the subtraction of the estimated nonselective component of transfer [Figure 9] from each curve of Figure 8 to yield the residual selective transfer. The symbols show observed values of residual selective transfer after various cue delays. The solid curves show the predictions that are based on vertical translations of a single generic selective-transfer curve [e.g., delay 0]. The form of the generic selective-transfer curve was estimated from the data.)

The assumption of additivity holds well for our data. The root-mean-square error is 0.016 for subject BL and 0.023 for subject BF.

Some Parametric Assumptions

The pure information transfer functions for the nonselective transfer process in Figure 9 and the selective transfer in Figure 10 can both be approximated by simple exponential growth functions of the form

$$f(t) = C[1 - \exp(-t/\tau)], \qquad (2)$$

where C is the asymptotic level of performance, and τ is the time constant of the growth process at an f(t) of 63% of C.

We denote nonselective transfer as $N_{c,m}(t)$, and selective transfer as $S_{c,m}(t)$. The indices remind us that transfer may depend not only on t, but also on the specific values of the cue delay, c, and the mask delay, m. For pure nonselective transfer, the cue comes after the mask, and we obtain

$$N_{c,m}(t) = C_N[1 - \exp(-t/\tau_N)], \qquad c \geq m, \qquad (3)$$

where C_N is the capacity of durable storage and τ_N is the time constant for nonselective transfer. Figure 9 shows Equation 3 with the constants chosen to optimize the fit to our data. The deviations of data from theory are very small.

Purely selective transfer occurs when the cue comes at (or before) the stimulus onset. Similarly to Equation 3, we obtain

$$S_{0,m}(t) = C_S[1 - \exp(-t/\tau_S)]. \qquad (4)$$

In Equation 4, C_S is the maximum number of letters the subject can transfer from one line. In general, C_S will be very close to the number of letters in one line. However, C_S has to be estimated because subjects are not perfect, and they occasionally miss a letter even in the easiest conditions. The time constant for selective transfer is τ_S.

Figure 10 shows the best fit of Equation 4 to our data for selective transfer. Again, deviations between the theory and the data are small. From Figures 9 and 10, we see that the growth rates of nonselective and selective information transfer curves are quite different, reflecting their different time constants, τ_N and τ_S.

Selective transfer for cue and mask delays with $c \leq m$ occurs only during the interval from c to m:

$$S_{c,m}(t) = S_{0,m}(t) - S_{0,c}(t). \qquad (5)$$

Of course, $S_{c,m}(t)$ is 0 whenever $c \geq m$. The total number of letters available for report in the cued row as a function of time is given by a generalization of Equation 1:

$$L_{c,m}(t) = S_{c,m}(t) + \tfrac{1}{3}N_{c,m}(t). \qquad (6)$$

The total number of letters available for whole reports is simply $N_{\infty,m}$.

A final complication is that the time a subject needs in order to interpret the cue may be greater than zero. To admit this possibility, a parameter τ_q, the cue interpretation time, is included in the model as an offset parameter, substituting $c + \tau_q$ for c in Equation 6.

Figure 12a summarizes the descriptive model. Two cumulative functions, $N_{c,m}(t)$ and $S_{c,m}(t)$, describe information transfer from iconic memory to durable storage. Before occurrence of a cue, transfer is governed by $N_{c,m}(t)$; after the cue, by $S_{c,m}(t)$. It is useful to think of the cue as a switch that toggles between the two transfer rates.

Predictions for a partial-report-plus-mask experiment are represented in Figure 12b. The number of letters available for report follows the trajectory to C_N until the occurrence of a cue. It then follows the trajectory described by $S_{c,m}(t)$. After the onset of the poststimulus mask, the predicted trajectory is flat.

Figure 12b shows that the cue is predicted to help most

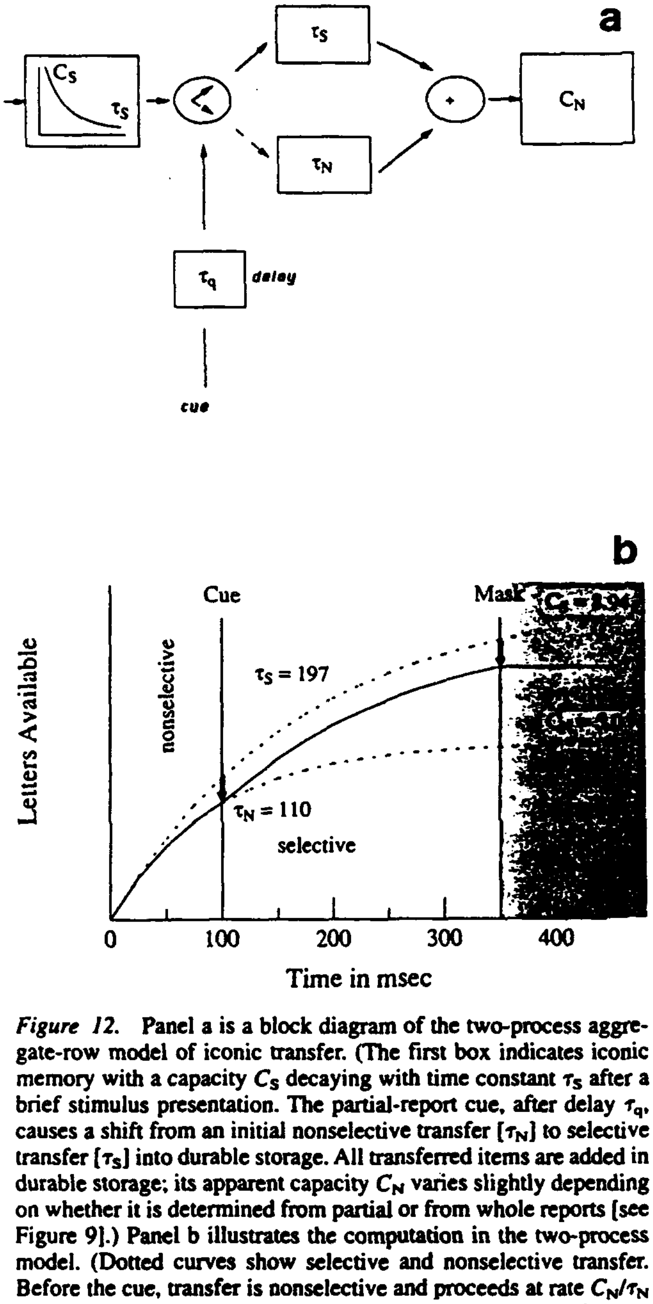


Figure 12. Panel a is a block diagram of the two-process aggregate-row model of iconic transfer. (The first box indicates iconic memory with a capacity C_S decaying with time constant τ_S after a brief stimulus presentation. The partial-report cue, after delay τ_q, causes a shift from an initial nonselective transfer [τ_N] to selective transfer [τ_S] into durable storage. All transferred items are added in durable storage; its apparent capacity C_N varies slightly depending on whether it is determined from partial or from whole reports [see Figure 9].) Panel b illustrates the computation in the two-process model. (Dotted curves show selective and nonselective transfer. Before the cue, transfer is nonselective and proceeds at rate C_N/τ_N to asymptote C_N. After the cue, transfer is selective at rate C_S/τ_S to asymptote C_S. The arrows indicate that in effect, the generic selective transfer curve is joined to the generic nonselective curve at the moment in time that the cue takes effect.)

when given within 100 ms of the stimulus onset. In the first 100 ms, the cumulative transfers $N_{c,m}(t)$ and $S_{c,m}(t)$ differ only slightly. After 100 ms, $N_{c,m}(t)$ reaches its asymptote, whereas $S_{c,m}(t)$ continues for at least another 300 ms.

Parameter Estimations and Fits to the Data

The curves in Figure 7 show the fit of the complete model to the data of both subjects. The model accounts for 96% and 97% of the variance in the data for Subjects BL and BF, respectively. The root-mean-square errors are 0.089 and 0.095, respectively.

The same subjects served in an earlier partial-report experiment, similar to Experiment 1 but without poststimulus masks. These earlier data can be fit by the model derived from Experiment 2 without estimating any new parameters. Figure 13 shows the partial-report-without-masking data and the model predictions. The parameter values were estimated from the experiment using masks. The fit obtained this way does not deviate significantly from the data. The dashed lines in Figure 13 show the contributions of nonselective transfer. The model indicates that, after stimulus termination, selective transfer decreases much faster than one might expect from the relatively slow decay in partial-report superiority.

Table 2 summarizes the parameter estimates for Subjects BF and BL. Two sets of estimates are shown. One set, already described, was derived from the subsets of the data that provided the pure nonselective transfer and the pure selective transfer analyses illustrated in Figures 9 and 10. A second set of parameters was estimated from the complete data of the partial-report-plus-masking experiments. The comparison of these estimates is an indicator of the overall consistency of the model, which is quite good.

For each subject, the nonselective capacity parameters, C_Ns, are very similar in the three relevant data sets: the full partial-report-plus-masking data set, the cue-after-noise subset, and the whole-report data set. The capacities are five letters for BL and seven letters (well above normal) for BF.

The nonselective capacity estimate C_N is effectively equal to 3, the number of letters in one row for BL. It is about 5–10% less than 4 for BF, who was shown four-letter rows.

The time constants for selective and nonselective transfer, τ_S and τ_N, are quite different from each other. Selective transfer continues to rise steadily until after 200 ms, whereas nonselective transfer asymptotes quickly after 100 ms. Both subjects have similar time constants, although their capacities differ.

For both subjects, the time τ_q necessary to interpret the cue is estimated to be 0 or slightly negative.

The speed of the transfer processes is determined by differentiating Equation 2. This results in

$$f'(t) = C/\tau \exp(-t/\tau). \quad (7)$$

For $t = 0$, Equation 7 reduces to

$$f'(0) = C/\tau. \quad (8)$$

Enigmas

Computation of the initial transfer rates immediately at stimulus onset, $S'(0)$ and $N'(0)$, shows that nonselective transfer has a much higher rate. Seventy and 45 letters/s are transferred nonselectively for subjects BF and BL, respectively, and only 27 and 15 are transferred selectively. Note that the nonselective transfer rates are based on the total number of letters transferred into durable storage, not merely on the letters in the cued row. In the aggregate model, it is not obvious why the actual speed of nonselective and se-

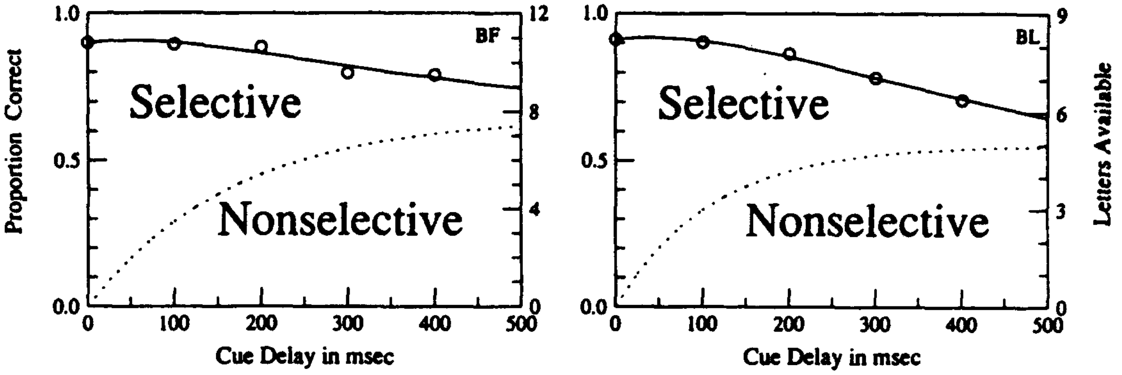


Figure 13. Data from a partial-report experiment without masks. (The open symbols show the proportion correct for various cue delays. The solid line shows the predictions of the two-process aggregate-row model. The dashed curve indicates the estimated component of performance resulting from nonselective transfer. The distance from the dashed line to the solid line [partial-report accuracy] represents the estimated contribution of the selective transfer process.)

lective transfer should appear to differ by so much; this issue is addressed in the row-by-row model, presented later.

It also is surprising that the estimated time the subject needs to interpret the cue τ_q is essentially zero. Experiments by Reeves and Sperling (1986) using visual cues showed that a spatial shift of visual attention took 300–400 ms. Sperling and Weichselgartner (in press) used a click in a go/no-go attention shift experiment that required only turning on attention, not actually shifting it in space. They found a modal switching time of about 100 ms. Our tonal cues, which required a three-choice reaction and a spatial shift of attention, would certainly be expected to have a much longer attention shift latency. These enigmas suggest the need for more complex analysis, which we provide in the next section by analyzing the data separately for each row.

Table 2
Best Fitting Parameter Values for the Aggregate-Row Model (Partial Report With Masking) and for Data Subsets That Yield Estimates of Pure Selective and Pure Nonselective Transfer in Experiment 2

Experiment	C_S	C_N	τ_S	τ_N	t_q	rms error
Subject BF						
Selective	3.58	—	130.0	—	—	0.073
Nonselective	—	7.08	—	99.9	−16.62	0.021
Combination	—	—	—	—	—	0.023
Overall	3.82	7.35	191.1	114.31	−12.5	0.095
Partial report	—	—	—	—	—	0.049
Whole report	—	6.3	—	—	—	—
Subject BL						
Selective	—	—	158.9	—	8.47	0.120
Nonselective	—	4.75	—	104.79	0.29	0.032
Combination	—	—	—	—	—	0.016
Overall	2.98	5.0	197.5	110.4	−13.6	0.089
Partial report	—	—	—	—	—	0.016
Whole report	—	5.0	—	—	—	—

Note. C_S and C_N represent attentional capacities, respectively, of selective and nonselective transfer, with units in letters; τ_s and τ_n represent time constants of selective and nonselective transfer, with units in milliseconds; rms is root mean square. Selective and nonselective experiments the parameters were estimated from subsets of the data that did not require using the additive combination rule. In the combination experiment, the combination rule that estimated only additivity, not any of the parameters, was listed. In the overall experiment, the complete model for Experiment 2 was tested. In the partial-report procedure, the partial-report-plus-mask parameters were used to predict the data from an earlier partial-report-without-mask experiment. The whole-report experiment entailed simply observation of subjects' performance. No parameters were estimated. A dash indicates that a parameter could not be estimated for a particular condition.

Position Effects

Partial-Report Accuracy by Row, Cue Delay, and Mask Onset Time

The probability of correct partial reports as a function of mask delay with cue delay as a parameter is displayed in Figure 14. Each panel shows data for a different row of the display. Partial reports of the middle row differ from reports of the top and bottom rows, and we consider the middle row first. Almost always, subjects report the middle row perfectly. Even in the hardest conditions (short mask delay and long cue delay) subjects report 80% of the middle-row letters correctly. Except for the earliest mask at 100 ms, all the other middle-row curves appear equal at a performance level of about 95% correct. There is no apparent iconic decay for the middle row. Obviously, the transfer to durable storage of letters from the middle row is nonselective, and in this the middle row differs from the other rows.

For the top and bottom rows, performance decreases from near perfect in easy conditions to near chance in the hardest conditions. Because nonselective transfer determines the asymptotic performance at long cue delays, the data for the top and bottom rows indicate there is much less nonselective transfer (and correspondingly more selective transfer) from these rows than from the middle row.

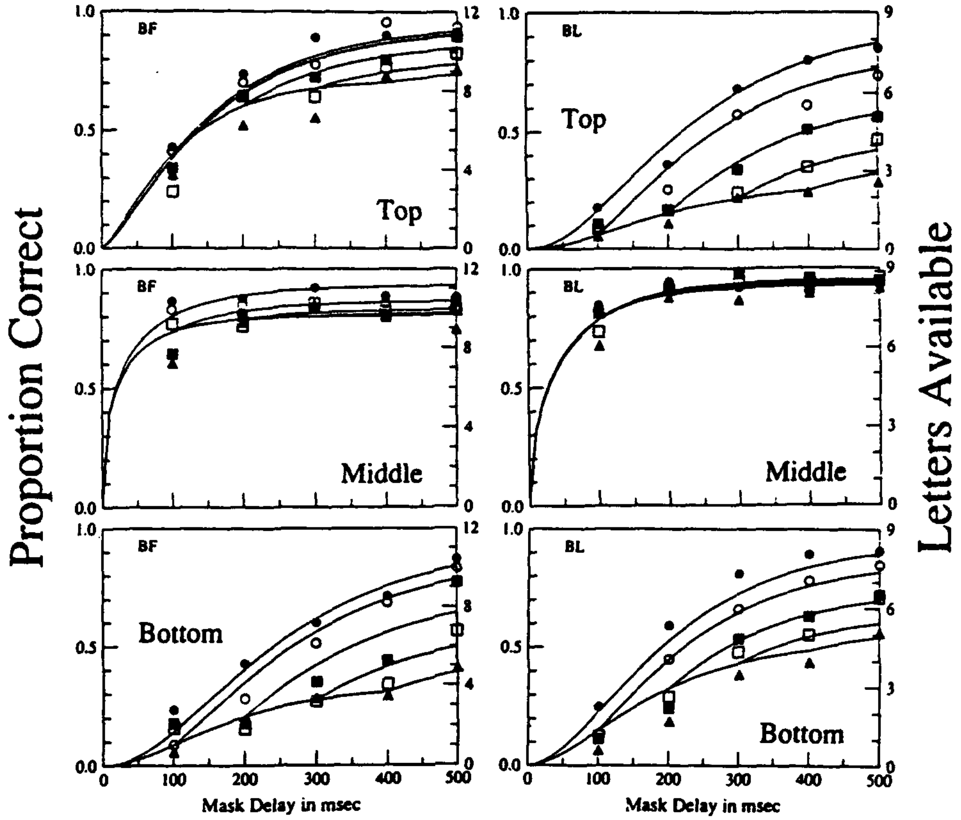


Figure 14. Accuracy of partial reports as a function of mask delay, with cue delay as a parameter, shown separately for each of the three stimulus rows. (Cue delays are in milliseconds: Filled circle = 0, open circle = 100, filled square = 200, open square = 300, triangle = 400. Top, Middle, and Bottom denote the stimulus rows. BL and BF denote the subjects. Each data point represents the proportion correct in 50–100 trials. Note the large and highly significant performance differences between the rows. The curves are predictions of the nine-parameter attentional model [Equations 12–14], with parameters given in Table 3.)

Selective and Nonselective Transfer by Row

To estimate the amount of nonselective transfer, we consider the subset of data with cue onset at or after mask onset (as in Figure 9). Figure 15 shows nonselective transfer for the three rows. For both subjects, the middle row rises to a high asymptotic level within the first 100 ms. The other rows rise slowly and reach generally lower asymptotic levels.

Nonselective transfer: Strategy mixture versus pure strategy. To account for the subjects' good performance on the middle row in the nonselective transfer data, we contrast two possibilities: a trial-to-trial variation of transfer strategy (which, over a series of trials, most often favors the middle row) and a consistent strategy that favors the middle row on every trial. Suppose that, prior to the stimulus exposure on each trial, subjects preselected a particular row to transfer nonselectively immediately following the exposure. Suppose that from trial to trial, they switched their preferred rows, but on the average, they most often chose the middle row. In this strategy, we would expect to find some trials for the top and bottom row on which the subjects' performances were perfect or nearly so. This is not the case, however. Of the trials on which the cue indicated a report of the top or bottom row, fewer than 2% of the reports had all letters correct (compared with 70% for the middle row). From this, we infer that subjects did not switch between rows and that the consistent preference of the middle row in nonselective transfer is responsible for its higher nonselective transfer rate. Indeed, most of the letters reported in the nonselective conditions come from the middle row.

On the other hand, in the conditions that favor selective transfer, the proportions with which the different rows are sampled are nearly the same. This explains why the aggregate model yielded faster rates of letters actually entering durable storage for nonselective than for selective transfer: Nonselective transfer sampled mostly the fast middle row, whereas selective transfer provided an almost equal mixture of all three rows.

Selective transfer. Figure 16 shows pure selective transfer estimated (as in Figure 10) from trials on which the cue occurred simultaneously with the onset of the test flash. For the middle row, selective transfer yields the

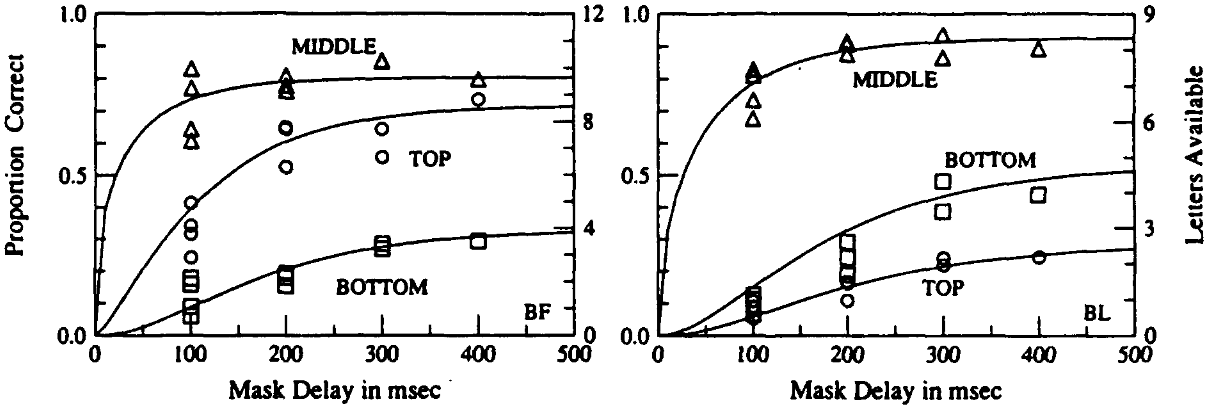


Figure 15. Pure nonselective transfer as a function of mask delay for each of three stimulus rows and 2 subjects. (The data points are the proportion of correct partial reports on trials in which the cue occurred at or after mask onset [Figure 6b]. Circles indicate the top row, triangles indicate the middle row, and squares indicate the bottom row. The solid curves are predictions of the nine-parameter attentional model [Equation 12], with parameters given in Table 3.)

same apparent growth curve as nonselective transfer. Note that, for both subjects, selective transfer for the top and bottom row reaches almost perfect performance at the longest mask delays. This indicates that the information in the stimulus is still available at these long delays. Therefore, it is also available for nonselective transfer. The finding that nonselective transfer reaches asymptote at a lower level for the top and bottom rows must then be a consequence of a capacity limitation. There is a suggestion in the data that the cumulative selective transfer from the top and the bottom rows is an S-shaped function of time. This would mean that the transfer rate was slow in the beginning, reached a maximum value at intermediate times, and finally declined again to zero. A slow start suggests a delayed shift of attention; the slow final rate almost certainly indicates that the iconic image has decayed to illegibility.

The additivity assumption of Equation 1 that yielded selective transfer by subtracting out the nonselective transfer can be applied to the row data to obtain the selective transfer for each individual row. The results are shown in Figure 17. We keep parallelism as a working hypothesis because it accounts for 99% of the variance of the data for both subjects.

Manifestations in the data of attention to a stimulus row. We assume that partial attention to a row slightly improves selective transfer of letters from that row relative to nonselective transfer and that complete attention to a row maximally facilitates selective transfer. Consider a graph of proportion correct versus mask delay with cue delay as the parameter (Figure 14). The earliest mask delay at which data from two cue delays, c_1 and c_2 diverge indicates the point at which the states of attention induced by c_1 and c_2 are sufficiently different to affect selective transfer. For example, consider cues that indicate the bottom row, and suppose $c_1 = 0$ and $c_2 = 100$ ms. In Figure 14, the data for $c_1 = 0$

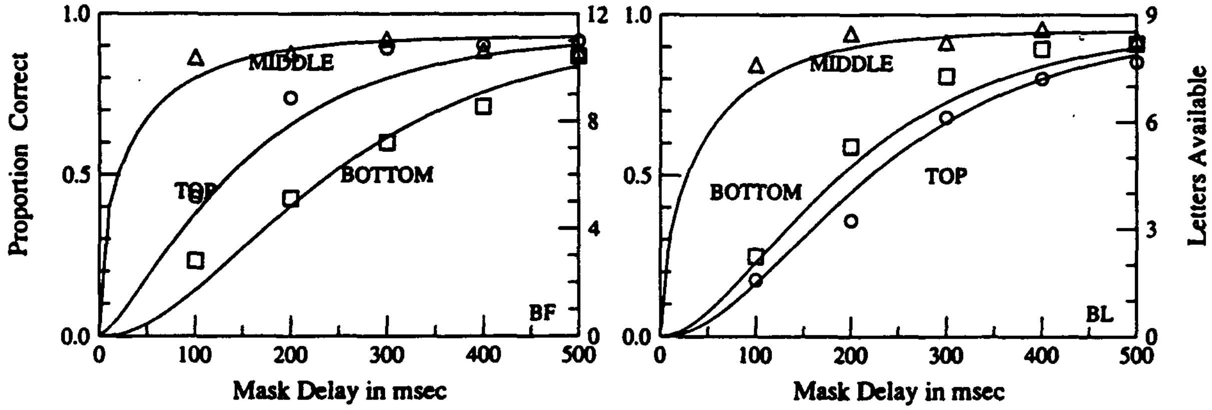


Figure 16. Pure selective transfer as a function of mask delay for each of three stimulus rows and 2 subjects. (Data points are the proportion of correct partial reports on trials in which the cue occurred at stimulus onset [Figure 6c]. Circles indicate the top row, triangles indicate the middle row, and squares indicate the bottom row. The solid curves are predictions of the nine-parameter attentional model [Equation 13], with parameters given in Table 3.)

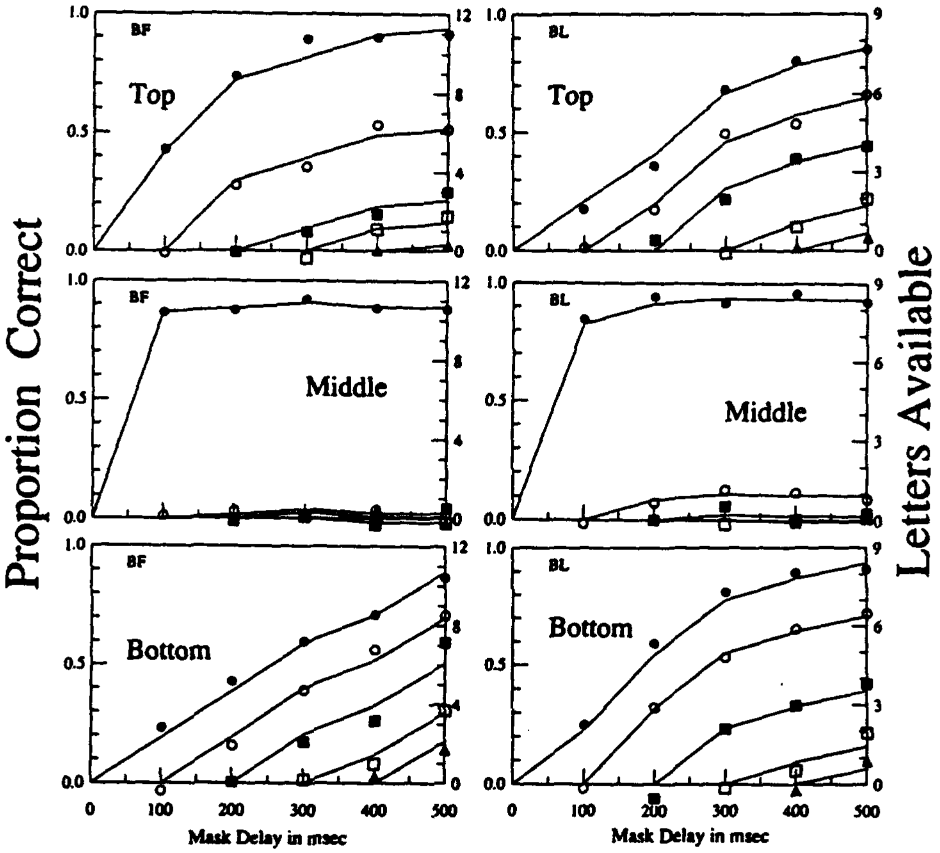


Figure 17. Selective transfer after prior nonselective transfer. (The curves are based on the same data as Figure 14—accuracy of partial reports as a function of mask delay, with cue delay as a parameter. Cue delays are in milliseconds: Filled circle = 0, open circle = 100, filled square = 200, open square = 300, triangle = 400. The estimated amount of nonselective transfer has been subtracted [as in Figure 11] to yield the residual selective transfer. The symbols show estimated values of residual selective transfer after various cue delays.)

first break away from the data for other cs when $m = 100$, and the $c_1 = 0$ data are completely separate when $m = 200$. A mask occurring 100 ms after c_1 means there is no further transfer from the stimulus after 100 ms. For the data obtained with c_1 in this condition to differ from the other c_i implies that the cue must have acted to alter attention within 100 ms. Alternatively, we would have to reject our previous assumption that the mask terminates stimulus availability.

Figure 14 shows that, for the middle row, there is no clear divergence of data for different cue delays and therefore no evidence that attention does or does not affect transfer of the middle row. However, transfer from the top and bottom rows is obviously quite affected by attention. The data for cue delay c in Figure 14 tends to break upward from the pack of longer cue delays as soon as $m \geq c$. This indicates that our cues induce a measurable change in attentional state immediately after their occurrence.

The other aspect of the performance-versus-mask-delay data (Figure 14) that we have already dwelt on at length is the parallelism of the curves for different cue delays onward from the moment $m \geq c$ (Figure 17). Parallelism indicates that the state of selective attention is the same for all the conditions represented in the parallel curve sections. In other words, not only does attention switch quickly once the cue arrives, but it switches completely. If it did not switch all at once, then an early cue, c_1, would have produced a greater attentional shift to the indicated row at a subsequent time, t_2 than a cue, c_2, that did not occur until t_2. In that case, transfer measured at t_2 would be faster for c_1 than for c_2, and the parallelism in the data of Figure 17 would be violated. Because the data are effectively parallel, we also have to assume that within the context of our assumptions, attention shifts quickly and completely. These assumptions are formalized in the next section.

Attentional Model of Transfer From Iconic Memory to Durable Storage

Assumptions

To account for the analysis of partial-report-plus-masking data separately by rows, we generalize the aggregate model of Figure 12a in a natural way, as illustrated in Figure 18. In

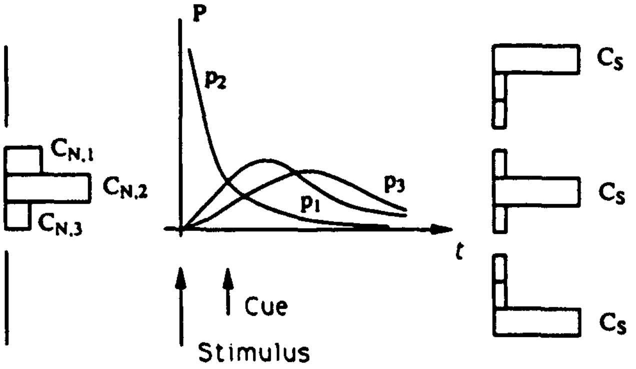


Figure 18. Illustration of the attentional model of iconic memory transfer processes. (As in the aggregate-row model, before the cue, the initial state of attention determines nonselective transfer from iconic memory to a durable store. $C_{N,1}$, $C_{N,2}$, and $C_{N,3}$, respectively, indicate the relative amounts of attention allocated to the top, middle, and bottom rows before the cue. The row-dependent retinotopic component of iconic transfer rate is illustrated by the p_r functions, which begin with stimulus onset. In response to a cue to report Row r, subjects shift attention instantaneously from its initial state to one of the three postcue states indicated by C_S. The actual transfer rate is the product of $C_{N,r}P_r(t)$ before the cue and $C_SP_r(t)$ afterward.)

the aggregate-row model, nonselective transfer and selective transfer were each characterized by a two parameters, their rate and the total capacity. Now these processes are made more explicit in terms of the states of attention they represent. Each state of visual attention is characterized by a spatial function that represents the allocation of attentional resources over space (Sperling & Weichselgartner, 1991) and a separable temporal function that represents the time period during which the spatial function reigns. Thus, nonselective transfer represents the default state of attention that exists from the beginning of the trial until the cue is ıeceived and interpreted. The spatial allocation of nonselective attention is described by three numbers ($C_{N,r}$, $r = 1, \ldots 3$) that represent the capacity (in letters) of durable storage allocated to the top, middle, and bottom rows of the display. The single number of the aggregate-row model that described the capacity of durable storage for nonselectively transferred letters was $C_N = \sum_r C_{N,r}$.

After the cue is received and interpreted, one of three states of selective attention occurs. Each state is characterized by C_s, the capacity allocated to the cued row (top, middle, or bottom) and by 0.0 capacity allocated to the other rows (Figure 18).

In the aggregate-row model, the two speeds of transfer from iconic memory to durable storage (nonselective and selective) were parameterized by their own time constants, τ_S and τ_N in Equation 7. In considering rows individually, it is obvious that all transfers are much quicker from the middle row and that three parameters, a_r, $r = 1, \ldots 3$, are needed to characterize the temporal differences between transfer rates from the three different rows. The exact temporal waveforms cannot be determined directly from our data. A mathematically tractable formulation that has useful properties is given by inserting a term $(t/\tau)^{a-1}$ to the time-dependent transfer of Equation 7, resulting in the time-and-row-dependent transfer

$$f'(t, a_r) = C_r(t/\tau)^{a_r-1}\exp(-t/\tau). \tag{9}$$

The constant C_r is a capacity allocated to row r.

We assume that the differences between nonselective and selective attention are completely captured by the spatial allocations of attentional capacity C_r, so that one set of row-dependent weights, a_r, suffices for all states of attention. Because the a_r depends on spatial location (the row) and does not depend on attention, it represents the intrinsic processing efficiency of a retinal location.

We wish to test our assumption that a single parameter suffices to describe the overall transfer rate of both selective and nonselective attention. Therefore, in parameter estimation, we estimate two overall rate parameters, one for selective and one for nonselective attention, to determine whether these unconstrained rate estimates indeed are similar.

Computational Model

Following Reeves and Sperling's (1986) attentional gating model, it is reasonable to assume that the transfer rate from a location, r, is determined by the product of two factors: (a) the availability (legibility) of stimulus information at r and (b) the amount of attention allocated to r. Availability at a location is determined by iconic buildup and decay, and it is parameterized by exponent, a_r, of Equation 9 combined with the exponential terms. Attentional allocation is parameterized by the capacity allocation, C_r. Equation 9 represents this product. Unfortunately, transfer mode appears implicitly in the time constant, τ, of Equation 9. This means that attention allocation (which determines transfer mode) would be inextricably intertwined with iconic availability if τ_N and τ_S were to differ appreciably.

Cumulative transfer in an interval [0, m] is given by integrating over Equation 9. For simplicity, we change the variable of integration, giving

$$S_{0,m}(m) = \int_0^{m/\tau} C_s(t)^{a_r-1}\exp(-t)\,dt. \tag{10}$$

By appropriately normalizing the exponential terms in Equation 10, we can convert Equation 10 into an attentional capacity (C, scaled in letters) multiplied by the well-known incomplete gamma function, $P(a, x)$, $0 \le P(a, x) < 1$:

$$P(a, x) = \frac{\int_0^x (t)^{a-1}\exp(-t)\,dt}{\int_0^\infty (t)^{a-1}\exp(-t)\,dt}. \tag{11}$$

In subsequent use, x will take four values: c/τ_N, c/τ_S, m/τ_N, and m/τ_S, representing the intervals from exposure onset to the cue and the mask, respectively, in units of the time constants of nonselective (τ_N) and selective (τ_S) transfer. For $a = 1$, Equation 9 simply reduces to Equation 7. Values of a between 0 and 1 lead to an accelerated exponential growth function for $P(a, x)$. Values higher than

1 lead to S-shaped delayed growth.

Equation 9 plays the same role in the row-by-row model as Equation 7 did in the aggregate-row model. It can be interpreted as representing the time course of the iconic image at each row location weighted by attentional allocation. To help the reader's intuition in following the exposition of the computational model, we show the iconic time course functions for three rows in Figure 18. The instantaneous transfer rate, $p_r(t) = t^{a-1} e^{-t}$, is the derivative of the cumulative transfer, $P(a, x)$, given in Equation 11. As Figure 18 indicates, the rise time for the middle row is too fast to be observed in our experimental conditions; however, the middle row decays with what appears to be a familiar, monotonically decreasing function. The top and bottom rows rise before they decay. We defer to later the question of whether these top- and bottom-row functions truly represent the rise and decay of iconic memory. (Alternatively, they might represent a property of a model in which the absence of independent measurements of attention insufficiently constrains the partition of transfer rates into legibility and attentional components.)

The following equations summarize the model. Equation 12 describes the cumulative nonselective transfer that takes place from the onset of the stimulus until either a cue or a mask occurs. It is the product of terms representing two factors: attentional allocation, C, and retinotopic/stimulus factors, P:

$$N_{r,c,m} = C_{N,r}P(a_r, c'/\tau_N), \qquad c' = \min(c, m). \quad (12)$$

The function P implies different transfer dynamics for each of the three rows r. Relative to the middle row, transfer from the top and middle rows is both delayed and slower. Because delay and slowness are perfectly correlated, both are captured by the parameter a_r.

There is a limit to the total number of letters that can be transferred from iconic to durable storage. How the subjects allocated space in durable storage to particular rows so as to optimize their performance is a matter that we did not attempt to control. Therefore, a parameter $C_{N,r}$ is needed for each row to describe the maximum number of letters of durable storage allotted to it (i.e., the default allocation of attention prior to the cue). Finally, the overall rate of nonselective transfer is determined by the time parameter τ_N.

Equation 13 describes selective transfer. In this formulation, selective transfer begins instantly at the onset of the cue and ends instantly at the onset of the mask:

$$S_{r,c,m} = C_S[P(a_r, m/\tau_S) - P(a_r, c/\tau_S)], \qquad m > c. \quad (13)$$

The cumulative transfer to durable storage depends on the integrated product of available information and attention (Reeves & Sperling, 1986). Available information is represented here by $P(a, x)$, which, for the special case of $\tau_N = \tau_S$ depends only on elapsed time since onset of the stimulus. Attention is represented by the currently operative set of C_r capacity values. Attention depends only on elapsed time since the onset of the cue. Therefore, when $\tau_N = \tau_S$, iconic time course and attention are independent.

Equation 14 expresses that the total number of letters transferred to durable storage from each row r is the sum of the nonselective and selectively transferred letters from r. It generalizes Equation 1 of the aggregate model:

$$L_{r,c,m} = N_{r,c,m} + S_{r,c,m}. \quad (14)$$

Parameter Estimates and Their Interpretation

Best fitting parameters were estimated for Equations 12–14 by means of an optimization program (PRAXIS; see Brent, 1973; Gegenfurtner, 1992), using the partial-report-plus-masking data of Experiment 2. The results of parameter estimation are summarized in Figure 14 and Table 3. The model's predictions correlate very well with the data: r^2 = .98 and .95 for the 2 subjects. The predicted average selective and nonselective transfer for the three rows is almost identical to the predictions of the aggregate-row model (see Figures 9 and 10). Therefore the row-by-row model (without additional parameters) also predicts the data from the whole-report experiment and from the partial-report experiment without masks.

The time constants τ_S for selective transfer and τ_N for nonselective transfer are now both approximately 100 ms, indicating that each transfer process completes in about the same time. However, the actual transfer rates C_r/τ depend on the row capacity. The aggregate-row model's capacity for nonselective transfer, C_N, is now split up into the three $C_{N,r}$s. This set of rates defines the initial default attention state prior to the cue. The high rate for the middle row indicates that the default attention state is primarily focused on the middle row. When a cue is received and interpreted, attention is shifted to the cued row, and the transfer rate is determined by the iconic legibility of that row, $f'(t, a_r)$ (Equation 9). In fact, the selective capacity, C_S, is virtually the same as in the aggregate-row model and nearly equal to the number of letters in the row. In effect, the model assumes that once attention is shifted away from the center row to the top or bottom row, it is as effective at the top or bottom row as it was in the center, and any difference in performance must be accounted for by differences in iconic legibility.

Parallel versus serial process in nonselective transfer.

Table 3
Best Fitting Parameter Values for the Model That Takes Differences Between Rows Into Account

Row	C_S	C_N	τ_S	τ_N	a
		Subject BF			
Top	3.72	2.87	115	82	1.39
Middle	3.72	3.21	115	82	0.38
Bottom	3.72	1.31	115	82	2.27
		Subject BL			
Top	2.85	0.85	109	97	2.25
Middle	2.85	2.78	109	97	0.50
Bottom	2.85	1.59	109	97	1.97

Note. C_S and C_N represent attentional capacities, respectively, of selective and nonselective transfer, with units in letters; τ_s and τ_n represent time constants of selective and nonselective transfer, with units in milliseconds; a is a pure number (Equation 12) that represents attentional dynamics.

The a_r parameters represent the dynamics of buildup and decay of iconic legibility at the retinal locations r. They represent the availability of information from a particular row regardless of whether it has been cued. In fact, some time after stimulus termination (100 ms for Subject BF, 200 ms for Subject BL; see Figures 15 and 17) the slopes of the nonselective transfer functions for the top and bottom rows in Figure 15 are still as steep or again become as steep as the initial slopes. This means that after 100 or 200 ms, the availability of information from these rows is as well as or better than it is immediately after stimulus termination.

One interpretation of the delayed availability of information from the top and bottom rows in nonselective transfer is that iconic legibility builds up slowly but approximately simultaneously in these noncentral rows. An alternative explanation is based on serial processes. Prior to a cue, subjects preprogram their attention to move away from fixation at about the time they expect to have completed transfer of the middle row to durable storage. Then they shift attention randomly to either the top or bottom row. This would result in an apparent delay in the availability of information from the top and bottom rows. If subjects were indeed shifting attention on nonselective report trials, it would greatly complicate the analysis of the attentive and iconic components of performance. The present data do not discriminate well between these alternatives.

Attention and the iconic time course are inextricably bound by multiplication in Equation 9: Only the product of attentional allocation and iconic availability determines performance. The model is a powerful computational device, but without an independent verification of the attentional state (or iconic availability), it is not a sufficiently precise tool to dissect unambiguously the attentional and iconic components of performance.

A comparison of the estimated values of a_1 and a_3 in Table 3 shows that the 2 subjects' iconic time courses are different in the top and bottom rows, with BF favoring the top row and BL the bottom row. Because these effects occur in both nonselective and selective attention, the model assigns them to the iconic time course. However, a more plausible interpretation would suggest that they represent tendencies, or biases, to shift attention up or down. According to this interpretation, in response to a cue, BF shifts his attention faster to the top row than the bottom row, and for BL it is just the reverse. These biases in selective transfer mirror the subjects' bias in nonselective transfer.

Although the model assumes that nonselective transfer reflects a single initial attentional state, closer examination of the data suggests that subjects first transfer letters from the middle row and then fill up the remainder of durable storage using nonselective transfer from the other rows. That is, even nonselective transfer ultimately may have to be modeled as consisting of two or more attentional states, an initial state of attention to the middle row, followed by attention to either the top or bottom row. Although this precision of description is necessary for the accurate partition of the components of performance (iconic decay, attention), it is not necessary from a purely computational point of view. The model accounts nicely for all the enigmas that remained after the aggregate-row model and provides a framework for dealing with the few problems that remain.

General Discussion

The present data show the critical importance of nonselective transfer in iconic memory experiments. By decomposing performance into selective and nonselective transfer and subtracting nonselective transfer from the total transfer, we were able to isolate the selective component that depends on the stimulus decay and attentional shifts. This isolation of the two transfer processes was made possible by using a completely crossed design of cue delay and mask delay. This crossed design differs from previous investigations with poststimulus masks (e.g., Averbach & Coriell, 1961; Irwin & Brown, 1987), in which only one mask or cue delay was used or cue and mask delay were correlated.

With respect to theory, we consider the three previous computational treatments of information transfer from iconic memory to durable storage. The earliest model (Averbach & Coriell, 1961) is extremely simple because it was developed for a more restricted paradigm. It proposes both a selective and a nonselective transfer process, but it embodies an assumption about probabilistic independence between these two processes that is strongly contradicted in our larger data set. Rumelhart's (1970) model is quite similar to ours. It fails because it embodies an incorrect assumption about subjects' strategies and another about memory capacity limits. These two models, and ours, share the common theme of two transfer processes. The third model (Loftus et al., 1985) derives iconic decay properties from a single nonselective transfer process. With respect to nonselective transfer and iconic decay, there is considerable agreement between Loftus et al.'s theory and ours, although their theory is not intended to confront the two transfer process issues that are our primary concern. In the next three sections, we consider these models in more detail. Then we briefly review noncomputational suggestions about iconic transfer processes.

Probabilistic Independence of Nonselective and Selective Transfer

Averbach and Coriell (1961) did a partial-report experiment in which the stimulus was two rows of 8 letters and the required partial report was a single letter. A visual cue ("bar marker") appeared above or below the required letter. Total transfer was determined in a partial-report experiment with a cue to report 1 of 16 possible letters. Nonselective transfer was estimated from report accuracy when the cued letter was masked with a concentric annulus. Selective transfer was estimated by correcting total transfer for the nonselective component. Because Averbach and Coriell's annulus was an effective letter masker only when the annulus occurred after a letter, and not when it occurred simultaneously, they ignored the data of the initial parts of their masking curves. However, their experimental results are generally similar to ours, even though the experimental conditions are quite different. Overall, performance was higher for our subjects.

Averbach and Coriell (1961) proposed the following combination rule. Their basic unit of analysis was a single letter, which could be transferred either selectively or nonselectively. They regarded the two transfer types as independent processes. Both transfers contribute probabilistically to the proportion of correctly reported letters, much as in Rumelhart's (1970) model. Averbach and Coriell found huge performance differences for different letter positions but decided to ignore them and average their data. Moreover, they did not vary mask and cue delays independently, so they were severely limited in what they were able to do with their data and theory. For example, they were noncommittal about whether nonselective transfer ends when the cue occurs, about whether selective transfer begins immediately upon cue onset, and about other issues related to the underlying processes.

.Averbach and Coriell's (1961) model is analyzed as follows. Each letter has a certain probability of being transferred by either process. Denote the event of a nonselective transfer with N, the event of a selective transfer with S, and the event of any kind of transfer with T:

$$P(T) = P(N) + (1 - P(N))P(S). \qquad (15)$$

It then follows that selective transfer is given by

$$P(S) = (P(T) - P(N))/(1 - P(N)). \qquad (16)$$

Equation 16 expresses the idea that two processes contribute to partial-report accuracy, as does our Equation 2 (and its subsequent elaborations) and Rumelhart's (1970) model (discussed shortly).

Figure 19 compares the rates of selective transfer (i.e., the legibility of the iconic image) as derived from the present model and from Averach and Coriell's (1961) model. It shows the number of letters selectively transferred during successive 100-ms intervals plotted as a function of the time at the end of the interval. The data points in Figure 19a are derived from Equation 2. Figure 11 shows cumulative selective transfer; Figure 19 shows selective transfer rate, that is, successive differences between the points of the lines in Figure 11. That all these successive differences fall on the same iconic decay function should be no surprise. We previously noted that all the curves of Figure 11 derive from a single generic selective transfer function.

Figure 19b shows the predictions for Averbach and Coriell's (1961) formulation (assuming that nonselective transfer stops after the occurrence of the cue). When the cue delay is zero, there is no nonselective transfer, and both our model and theirs give the same predictions (indicated by the filled circles). However, when nonselective and selective transfer are combined (i.e., for any cue delay greater than zero), the models differ. As already pointed out in the discussion of Figure 11, our assumptions result in selective transfer rates that depend only on the time since the onset of the cue. Averbach and Coriell's model leads to large, highly irregular estimates of selective transfer for a given time interval, and no clear pattern emerges of how selective transfer is determined. Their model cannot account for our data.

The consistency of the different independent estimates of the iconic decay function demonstrates the value of our Cue Delay × Masking Delay crossed design, which enables us not only to estimate the parameters for our model, but also to check our model's consistency.

Diffuse Transfer Followed by Focused Transfer

Rumelhart (1970) proposed a mathematical model of partial-report experiments cast in terms of features. Features were transferred with replacement from retinal locations and aggregated to form letters. During the stimulus exposure, features were equally available at all locations and all times. After termination of the exposure, feature availability was assumed to decay exponentially. The feature extraction rate was assumed to have an absolute limit (capacity). Before a cue was received, the overall feature extraction capacity was spread equally over all locations. Immediately after a cue was received, feature extraction capacity was concentrated entirely on the cued locations.

The essential ideas of Rumelhart's (1970) model are quite similar to those of our model, namely, that there is a default precue attentional state followed by a postcue attentional state and that the same transfer process operates in both states (merely the row allocations are different). However, Rumelhart was unaware that the precue state is not diffusely spread over all rows but is concentrated on the middle row. In addition, he had no explicit capacity limit for durable storage, relying on limited stimulus availability to account for all response limitations. This was obviously too restrictive an assumption.

Rumelhart's (1970) representation of the probability of correct reports, P, as the indirect result of a feature extraction process would allow the P versus time graphs either to grow like exponentially limited growth processes or to assume S shapes. An S shape would result from the fact that before a threshold number of features is collected at a location, the probability of correctly reporting the letter at that location is assumed to be at chance. After the critical number of features is collected, the probability of correct report is assumed to be 1.0. Although there is considerable flexibility in the generation of S-shaped curves under the feature accumulation assumption, and our empirical P versus t curves are, in a few cases, S shaped, it seemed better not to burden our transfer theory with such a complex assumption.

How Much Is an Icon Worth?

The nonselective transfer curves obtained in our experiments appear very similar to the ones derived by Loftus et al. (1985) in a paradigm using pictorial stimuli. They measured the number of details subjects could report from briefly exposed pictures. Exposure duration was varied, and a mask followed stimulus presentation immediately after stimulus offset. In a second condition, presentation of the mask was delayed 300 ms. They found that a 300-ms mask delay after

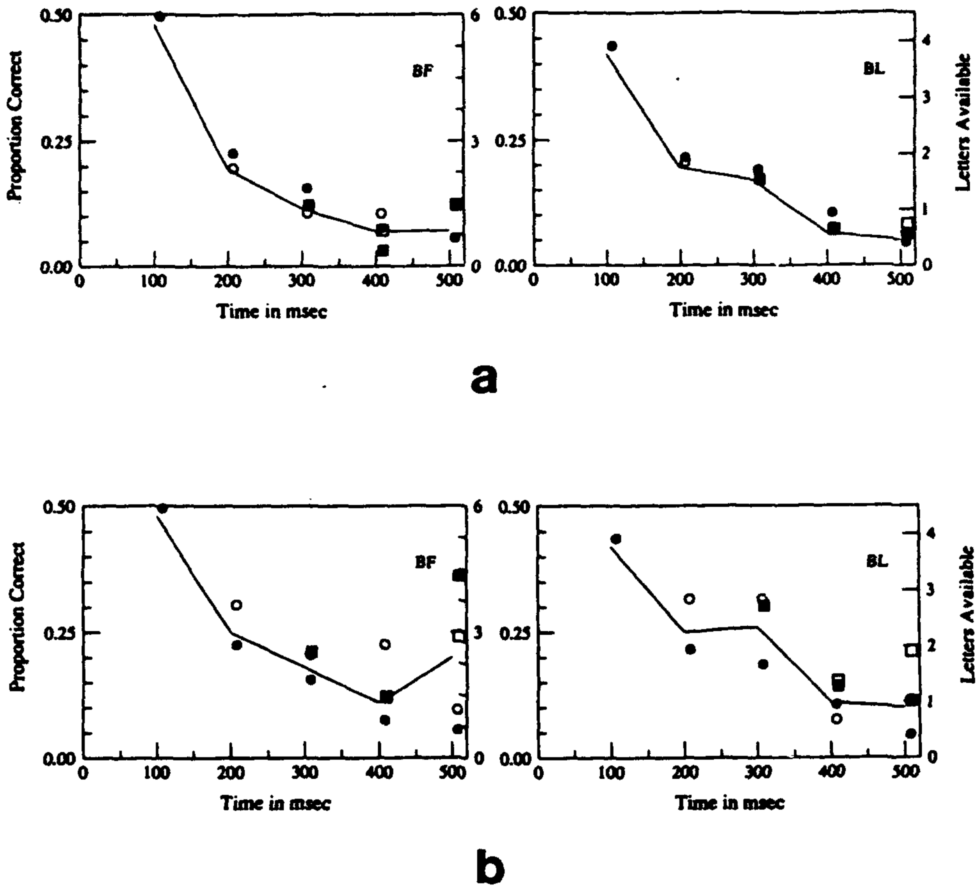


Figure 19. Derived iconic memory decay functions: Estimates of the rate of selective transfer at a given time after onset of the 50-ms stimulus. (At each time *t*, selective transfer during the 100 ms preceding *t* is estimated independently from each condition, with cue delay less than *t*. [This requires extraction of the selective transfer component from total transfer whenever cue delays are greater than 0.] Symbols indicate cue delays in milliseconds: Filled circle = 0, open circle = 100, filled square = 200, open square = 300, triangle = 400. Data are shown for Subjects BF [left] and BL [right]. Panel a shows estimates of selective transfer derived from our model [Equation 5]. Data points are the differences between successive points on each of the lines of Figure 11. Insofar as the different estimates all fall on the same iconic decay function, it substantiates our model of iconic decay. Panel b shows estimates of selective transfer derived from Averbach and Coriell's [1961] model [Equation 16] applied to our data. The wide variation [at a given time] of the different estimates for selective transfer indicates that this model does not yield a consistent description of iconic decay.)

the termination of a stimulus exposure led to the same levels of performance as an additional 100-ms stimulus exposure. They argued that an additional exposure of 100 ms is equivalent to an icon that is available for 300 ms. In this and subsequent experiments (Loftus, Duncan, & Gehrig, 1992; Loftus & Hogden, 1988), with various stimulus materials and tasks, they equilibrated iconic availability against an equivalent continued exposure that yielded the same performance. Ultimately, Loftus et al. (1992) derived an iconic decay function in terms of equivalent continuation of the stimulus. Their derived iconic decay functions were approximately, but not precisely, exponential.

The assumptions underlying Loftus et al.'s (1992) and our analyses are quite similar, although they derived all their data from whole reports. The main difference is that we present iconic decay directly in terms of a transfer rate; whereas they presented it in terms of the fraction of the transfer rate of a continued stimulus exposure. Furthermore, in their procedures, it apparently was not necessary to discriminate the transfer rate at different retinal locations, which is critical in our analyses. Their derived iconic decay functions agree quite well with those we derive for the middle row.

Position Effects

Differences in performance for different parts of the display have long been observed (Averbach & Coriell, 1961; Holding, 1970; Sperling, 1960); but have not been taken into account in the estimation of the duration of iconic memory. The differences we observe are mainly lower transfer rates for the top and bottoms compared with the middle row. Although in our formulation these locational factors are tied to the stimulus, it is likely that they are based, at least in part, on attentional factors. A relevant positional analysis was done by Holding (1970). He varied the probability with which each row was cued and found that performance varied accordingly, implicating attention. However, Holding's analysis was insufficient to discriminate a change in precue nonselective strategy from (postcue) difference in iconic decay. Furthermore, like many others (see Long, 1980), we strongly disagree with Holding's conclusion that this observation can explain partial-report superiority without postulating an intermediate store.

When a rapidly moving spot is illuminated by stroboscopic flashes (temporally sampled motion), more than one spot appears to move simultaneously. The number of apparently visible spots is a measure of visible persistence, and this number varies with retinal location, persistence being longer for peripheral than for foveal stimuli (Farrell, Pavel, & Sperling, 1990). Eventually, such spatial nonhomogeneities of the visual system will have to be reflected in accounts of iconic decay.

Strategy

The results obtained in Experiment 1 seem to contradict earlier results by Sperling (1960) that showed an influence of subjects' strategy. In Experiment 1, performance for a given cue delay varied depending on which cue delays were given in preceding sessions. We can resolve this contradiction by looking at subjects' overall performance level. Our subjects were well practiced. Under ideal conditions (no mask or cue delay), they achieved a performance level of 95–100% correct. Subjects in Sperling's study achieved 70–95% correct under the same conditions. This suggests that these early strategy effects were due to the use of nonoptimal strategies that would have been discarded after additional practice.

Other Models

In recent studies, Irwin and Brown (1987) and Irwin and Yeomans (1986) tested an alternative conception of iconic memory. Their theory also has two buffers, analogous to the iconic memory and short-term memory concepts of traditional theories. It assumes that the coding in iconic memory has two separate representations, one for identity and another for spatial location. This distinction does not bear directly on the distinction between nonselective and selective transfer, but it is relevant to the general issue.

We did not analyze our data for errors of intrusion and location, but we have some important relevant observations from serving as subjects ourselves and from speaking to subjects in iconic memory experiments. When a subject happens to be attending to one row while another one is cued (and fails to perceive the cued letters), the tendency is not to guess at random, but to report the (nonselectively) transferred letters even though they are known to be in the wrong row. To a subject, it seems better to report letters at least known to have been somewhere in the stimulus than to report random letters; the reasoning perhaps being that the cue or the rows may have been misperceived. Therefore, in assessing location errors, it is critical to use additional measures to assess the nature of the errors. For example, Sperling and Dosher (1986) noted that when items were reported with high confidence, location errors were extremely rare and practically never extended beyond an adjacent location. Irwin and Yeomans (1986) supported the notion of row juxtaposition. In their study, incorrect letters mostly came from an incorrect row in the correct column of the display.

Summary and Conclusion

We experimentally identified two transfer processes, nonselective and selective, in the partial-report task. Our data provided strong evidence that performance in the partial-report task is given by the algebraic sum of these two processes. Experiment 1 showed that independent of cue delay, subjects use only one strategy in a partial-report experiment. Experiment 2 showed that this strategy consists of nonselectively transferring letters until the cue appears and afterwards selectively transferring them.

The many complexities of these experiments are accurately described by a computational model that makes several plausible assumptions. Transfer rates are determined by the product of iconic legibility of the stimulus (which depends on the elapsed time after stimulus exposure and on the retinal location) and the subject's attentional state. Nonselective transfer is characterized by rapid transfer of the middle row and much slower transfer of other rows. This precue attentional state is parameterized in the computational model by the precue capacity allocations weighted heavily toward the middle row.

Immediately after the cue, attention shifts to the cued row of the display. Postcue capacity allocation is maximum for the cued row and zero for the others. From this moment on, until the poststimulus mask ends all iconic transfer, selective transfer occurs from the cued row. Nonselective transfer is focused mainly on the middle row, whereas selective transfer focuses exclusively on the cued row, so that selective transfer produces more correct items on the average—a higher effective transfer rate. However, empirically determined rate constants (completion times) for nonselective and selective transfers are approximately the same ($\tau \approx 100$ ms), suggesting that all transfers represent the same process and the different effective rates reflect different states of attention, different retinal locations, and different likelihoods that the transferred items will be in the cued row.

References

Adelson, E. H., & Jonides, J. (1980). The psychophysics of iconic storage. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 486–493.

Averbach, E., & Coriell, E. (1961). Short-term memory in vision.

Bell System Technical Journal, 40, 309–328.

Averbach, E., & Sperling, G. (1960). Short term storage of information in vision. In C. Cherry (Ed.), *Information theory* (pp. 196–211). London: Butterworth.

Brent, R. P. (1973). *Algorithms for minimization without derivatives.* Englewood Cliffs, NJ: Prentice Hall.

Budiansky, J., & Sperling, G. (1969). *GSLetters. A general purpose system for producing visual displays in real time and for running psychological experiments of the DDP24 computer* (Bell Telephone Laboratories Technical Memorandum 69-1223-6). Murray Hill, NJ: AT&T Bell Laboratories.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228.

Dick, A. O. (1969). Relations between the sensory register and short-term storage in tachistoscopic recognition. *Journal of Experimental Psychology, 82,* 279–284.

DiLollo, V. (1984). On the relationship between stimulus intensity and duration of visible persistence. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 144–151.

Duncan, J. (1983). Perceptual selection based on alphanumeric class: Evidence from partial reports. *Perception & Psychophysics, 33,* 533–547.

Efron, R. (1970). Effect of stimulus duration on perceptual onset and offset latencies. *Perception & Psychophysics, 8,* 231–234.

Eriksen, C. W., & Collins, J. F. (1967). Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology, 74,* 476–484.

Farrell, J. E., Pavel, M., & Sperling, G. (1990). The visible persistence of stimuli in stroboscopic motion. *Vision Research, 30,* 921–936.

Gegenfurtner, K. R. (1992). PRAXIS: Brent's algorithm for function minimization. *Behavior Research Methods, Instruments, & Computers, 24,* 560–564.

Hall, D. C. (1974). Eye movements in scanning iconic imagery. *Journal of Experimental Psychology, 103,* 825–830.

Hogden, J. H., & Di Lollo, V. (1974). Perceptual integration and perceptual segregation of brief visual stimuli. *Vision Research, 14,* 1059–1069.

Holding, D. H. (1970). Guessing behavior and the Sperling store. *Quarterly Journal of Experimental Psychology, 22,* 248–256.

Irwin, D. E., & Brown, J. S. (1987). Tests of a model of informational persistence. *Canadian Journal of Psychology, 41,* 317–338.

Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 343–360.

Kaufman, J. (1978). *Visual repetition detection.* Unpublished doctoral dissertation, New York University, New York.

Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin, 70,* 404–425.

Kropfl, W. J. (1975). *Variable raster and vector display processor* (Bell Telephone Laboratories Technical Memorandum 75-1223-3). Murray Hill, NJ: AT&T Bell Laboratories.

Külpe, O. (1904). Versuche über Abstraktion [Experiments on abstraction]. In V. F. Schumann (Ed.), *Bericht über den 7. Kongress für experimentelle Psychologie* (pp. 56–68). Leipzig: Barth.

Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 530–549.

Loftus, G. R., & Hogden, J. (1985). Extraction of information from complex visual stimuli: Memory performance and phenomenological appearance. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 139–191). San Diego, CA: Academic Press.

Loftus, G. R., Johnson, C. A., & Shimamura, A. P. (1985). How much is an icon worth? *Journal of Experimental Psychology: Human Perception and Performance, 11,* 1–13.

Long, G. M. (1980). Iconic memory: A review and critique of the study of short-term visual storage. *Psychological Bulletin, 88,* 785–820.

Melchner, M. J., & Sperling, G. (1980). *VEX: A computer system for real-time vision experiments* (Bell Telephone Laboratories Technical Memorandum 80-1223-4). Murray Hill, NJ: AT&T Bell Laboratories.

Merikle, P. M., Lowe, D. G., & Coltheart, M. (1971). Familiarity and method of report as determinants of tachistoscopic performance. *Canadian Journal of Psychology, 25,* 167–174.

Mewhort, D. J. K., Campbell, A. J., Marchetti, F. M., & Campbell, J. I. D. (1981). Identification, localization, and "iconic memory": An evaluation of the bar-probe task. *Memory & Cognition, 9,* 50–67.

Mewhort, D. J. K., Johns, E. E., & Coble, S. (1991). Early and late selection in partial report: Evidence from degraded displays. *Perception & Psychophysics, 50,* 258–266.

Mewhort, D. J. K., Merikle, P. M., & Bryden, M. P. (1969). On the transfer from iconic to short-term memory. *Journal of Experimental Psychology, 81,* 89–94.

Neisser, U. (1967). *Cognitive psychology.* Englewood Cliffs, NJ: Prentice Hall.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. J. Pick & I. J. Saltzman (Eds.), *Modes of perception* (pp. 137–157). Hillsdale, NJ: Erlbaum.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93,* 180–206.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology, 7,* 191–216.

Sakitt, B. (1976). Iconic memory. *Psychological Review, 83,* 257–276.

Scarborough, D. L. (1972). Memory for brief visual displays of symbols. *Cognitive Psychology, 3,* 408–429.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* 1–29.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 19–31.

Sperling, G. (1967). Successive approximations to a model for short-term memory. *Acta Psychologica, 27,* 285–292.

Sperling, G., & Dosher, B. (1986). Strategy and optimization in human information processing. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and performance* (Vol 1, pp. 1–65). New York: Wiley.

Sperling, G., & Gegenfurtner, K. (1988). Two transfer processes in iconic memory. *Bulletin of the Psychonomic Society, 26,* 488.

Sperling, G., & Weichselgartner, E. (in press). Episodic theory of the dynamics of spatial attention.

Townsend, V. M. (1973). Loss of spatial and identity information following a tachistoscopic exposure. *Journal of Experimental Psychology, 98,* 113–118.

Wundt, W. (1899). Zur Kritik tachistoskopischer Versuche [A criticism of tachistoscopic experiments]. *Philosophische Studien, 15,* 287–317.

Received March 28, 1991
Revision received September 30, 1992
Accepted October 6, 1992 ■

Vision Res. Vol. 33, No. 4, pp. 463–485, 1993
Printed in Great Britain. All rights reserved

0042-6989/93 $6.00 + 0.00

The Dimensionality of Texture-Defined Motion: a Single Channel Theory

PETER WERKHOVEN,*† GEORGE SPERLING,* CHARLES CHUBB*‡

Received 22 August 1991; in revised form 3 August 1992

We examine apparent motion carried by textural properties. The texture stimuli consist of a sequence of grating patches of various spatial frequencies and amplitudes. Phases are randomized between frames to insure that first-order motion mechanisms *directly* applied to stimulus *luminance* are not systematically engaged. We use ambiguous apparent motion displays in which a heterogeneous motion path defined by alternating patches of texture s (standard) and texture v (variable) competes with a homogeneous motion path defined solely by patches of texture s. Our results support a one-dimensional (single-channel) model of motion-from-texture in which motion strength is computed from a single spatial transformation of the stimulus—an *activity* transformation. The value assigned to a point in space–time by this activity transformation is directly proportional to the modulation amplitude of the local texture and inversely proportional to local spatial frequency (within the range of spatial frequencies examined). The activity transformation is modeled as the rectified output of a low-pass spatial filter applied to stimulus *contrast*. Our data further suggest that the *strength* of texture-defined motion between a patch of texture s and a patch of texture v is proportional to the product of the activities of s and v. A strongly counterintuitive prediction of this model borne out in our data is that motion between patches of different texture can be stronger than motion between patches of similar texture (e.g. motion between patches of a low contrast, low frequency texture l and patches of high contrast, high frequency texture h can be stronger than motion between patches of similar texture h).

Second-order motion Motion metamers Motion energy Motion correspondence

INTRODUCTION

First-order motion extraction

Drifting spatiotemporal modulations of various sorts of optical stuff (such as luminance, contrast, texture, binocular disparity, etc.) can induce vivid motion percepts; in each case "something" appears to move from one place to another. This introspective description, however, does not necessarily reflect the underlying processes in human visual motion processing.

The study of visual motion extraction mechanisms has traditionally focused on rigidly moving objects, projecting drifting modulations of *luminance*. Several physiologically plausible computational models have been proposed to extract motion information from drifting luminance modulations. Examples are the gradient detector (see Moulden & Begg, 1986) and the Reichardt or correlator detector (see Reichardt, 1961). These detectors are designed to detect drifting luminance modulations (or their linear transformations) and are therefore called *first-order* motion extraction mechanisms (Cavanagh & Mather, 1989)

Psychophysical experiments (e.g. van Santen & Sperling, 1984; Werkhoven, Snippe & Koenderink, 1990b) have shown that motion perception of drifting modulations of luminance is well explained by a first-order computation called *motion energy extraction*. Indeed, most current models of first-order motion detection (e.g. Reichardt detectors and gradient detectors) have now been shown to be equivalent or approximately equivalent to some variant of motion energy extraction (Adelson & Bergen, 1986; van Santen & Sperling, 1985). A standard approach to first-order motion energy extraction (e.g. Heeger, 1987; Adelson & Bergen, 1985) proposes that the visual system uses a battery of spatiotemporally oriented filters, each of which yields a real-valued function of the visual field over time. The output of each filter is squared at each location in space to obtain a measure of local energy at the spatiotemporal frequency to which that filter is tuned. The squared outputs of these filters (motion energies) comprise the input to a higher order process that computes a velocity flow field. For example, Heeger's (1987) model is built on the observation that the Fourier transform of a rigidly translating pattern has all its energy contained in a plane through the origin in frequency space. Each motion energy detector (narrow-band, spatiotemporal linear

*Department of Psychology and Center for Neural Science, New York University, New York, NY 10003, U.S.A.

†Present address: Utrecht Biophysics Research Institute (UBI), Buys Ballot Laboratory, Utrecht University, Princetonplein 5, 3584 CC, Utrecht, The Netherlands.

‡Present address: Department of Psychology, Rutgers University, New Brunswick, NJ 08903, U.S.A.

filter followed by squaring) has its energy confined to a Gaussian neighborhood of frequency space near the origin. The velocity vector assigned a given point in space at a given time is obtained by (i) weighting the energy spectrum of each detector by that detector's response, and (ii) finding the plane through the origin of frequency space that absorbs the greatest amount of this locally measured motion energy.

Second-order motion extraction

Chubb and Sperling (1988, 1989a, b, 1991) demonstrated broad classes of *drift-balanced* and *microbalanced* stimuli that clearly appeared to move but for which even complete knowledge of the energy of all their Fourier components would be useless in deciding whether their motion was to the left or to the right (see also Cavanagh, Arguin & von Grünau, 1989; Lelkens & Koenderink, 1984; Mather, 1991; Ramachandran, Rau & Vidyasagar, 1973; Turano & Pantle, 1989; Victor & Conte, 1990). Thus first-order motion energy extraction fails completely to account for the perception of motion in drift-balanced stimuli. Such stimuli are said to elicit *second-order motion perception* (Cavanagh & Mather, 1989; Chubb & Sperling, 1988). In second-order motion stimuli, what drifts is not a luminance modulation but modulation of contrast, or spatial frequency, texture type, flicker, or some other stimulus property.

Stages. Let L be the spatiotemporal luminance function defining a stimulus. The luminance at point (x, y) at time t is then denoted $L(x, y, t)$. In our analysis, we discriminate three stages for the extraction of motion information from L: preprocessing; flow field extraction; and decision.

First, a preprocessing stage in which one or more transformations T_i are applied to L yielding a set of real-valued, time-varying, "neural images" $T_i(L)$ (Robson, 1980). The value at point (x, y) at time t that results from applying T_i to L is thus denoted $T_i(L)(x, y, t)$. Usually, we think of i as referring to the dominant spatial frequency of a transformation—its scale.

Second, each time-varying, neural image $T_i(L)$ is the input to a motion-analysis stage $\vec{V}_i$ whose output is a (time-varying) velocity flow field $\vec{V}_i \bigcirc T_i(L) = \vec{V}_i[T_i(L)]$. For any point (x, y) in the visual field and every time t, the value $\vec{V}_i \bigcirc T_i(L)(x, y, t)$ is a two-dimensional vector that indicates estimated pattern velocity of the transformed image $T_i(L)$ in the neighborhood of (x, y) at time t. The scale of $\vec{V}_i$ corresponds to T_i. Associated with $\vec{V}_i \bigcirc T_i(L)$ is a real-valued function $S_i(L)$ that gauges the reliability or strength of the velocity estimate provided by $\vec{V}_i \bigcirc T_i(L)$. For instance, the velocity estimate obtained at point (x, y) and time t may have been computed from sparse or noisy data. In this case, irrespective of estimated direction or speed, the strength $S_i(L)(x, y, t)$ of the estimated velocity $\vec{V}_i \bigcirc T_i(L)(x, y, t)$ may be low.

Finally, all the velocity flow fields $\vec{V}_i \bigcirc T_i(L)$ and their associated strength maps $S_i(L)$ feed into a decision mechanism; its output determines the direction of apparent motion in ambiguous displays.

The preprocessing transformation T_i can be either linear or nonlinear. Generalizing previous terminology, we say that any system that employs linear preprocessing performs first-order motion extraction, whereas nonlinear preprocessing performs second-order motion extraction (e.g. Cavanagn *et al.*, 1989; Chubb & Sperling, 1988).

We refer to the transformations $\vec{V}_i \bigcirc T_i$ as motion channels. T_i is called the initial transformation and $\vec{V}_i$ the motion extractor. S_i is called the strength measure of the channel.

Motion-energy detection vs motion-correspondence detection. Both first- and second-order motion channels can be further classified by the type of motion extraction they use. A review of the literature on motion perception shows that two types of motion extractor have been considered and tested experimentally. We call these types of motion extraction motion energy extraction and motion correspondence extraction.

Motion *energy* extraction computes the directional energy of a Fourier representation of the drifting modulation signal, that is, the relative energy of "drifting" spectral components. Within the constraints set by frequency resolution, energy extraction is independent of the relative phase of the different spatial Fourier components of the modulation signal (van Santen & Sperling, 1984). In this respect, motion energy extraction computations are largely insensitive to similarities between items in a motion path. The first-order motion analysis models noted above (Reichardt, 1961; Adelson & Bergen, 1985; Marr & Ullman, 1981) all share this property.

Traditionally, however, psychophysicists have interpreted results of a wide range of motion experiments in terms of correspondence extraction. The metaphor of correspondence extraction describes motion as the convection of some invariant aspects of spatial structure over time. Thus, motion correspondence extraction depends on similarity of local features. The more nearly similar are two adjacent features that are separated by an interval in time, the greater will be the strength of motion between them.

The distinction between motion energy extraction and motion correspondence extraction can be summarized as follows: let α and β be two points separated by a brief interval in space and time, and let v_α and v_β be the stimulus intensities at α and β. Then motion energy extraction yields a motion strength that is a monotonically increasing function of the product $v_\alpha v_\beta$. Motion correspondence extraction yields a motion strength between α and β that is a decreasing, nonnegative function of $|v_\alpha - v_\beta|$.

Typically, motion channels using correspondence extraction yield higher motion strengths between similar textures than between dissimilar textures. In particular, a motion channel using a correspondence extractor can never yield motion strength between a patch of optical stuff A and a patch of different stuff B that is greater than the motion strength between two patches of stuff A. This can easily happen, however, for motion channels using energy extractors. Suppose, for instance, that $v_A > v_B$, for v_A and v_B the respective values assigned stuff A and stuff B by a channel's initial transformation. Then, motion energy extraction yields greater strength of motion between a patch of A and a patch of B ($v_A v_B$) than between two patches of B ($v_B v_B$).

Motion-from-texture

The purpose of this paper is to characterize the mechanism of second-order motion perception in the subclass of drift-balanced stimuli for which motion is defined by a modulation of spatial texture properties. To reiterate, it is not produced by a moving texture patch—that would be rigid, luminance-defined motion. Texture-defined motion is most conveniently produced by a moving patch that is filled with a particular type of texture in which each successive frame represents a new, uncorrelated instance of that texture type (Chubb & Sperling, 1989a, 1991). As is true for all drift-balanced motion stimuli, an intriguing aspect of texture-defined motion perception is that (unlike perception of luminance defined or first-order motion) it cannot be explained by Fourier energy or autocorrelational motion analysis (standard motion analysis).

An early example of texture-defined motion was reported by Ramachandran *et al.* (1973). Detailed studies and analysis were recently presented by Chubb and Sperling (1988, 1989a, b, 1991), Cavanagh *et al.* (1989), Mather (1991), Turano and Pantle (1989), and Victor and Conte (1989).

We construct stimuli for which energy and correspondence mechanisms yield different predictions for the strength of texture-defined motion (Werkhoven *et al.*, 1990b). The resulting data demonstrate that texture-defined motion is computed by an energy mechanism, and not a correspondence mechanism. And we will show how psychophysical data can be used to discriminate between these two sorts of mechanisms in human perception of texture-defined motion. More importantly, these data indicate clearly that, for the class of textures we use (similarly oriented patches of random-phased sinusoidal grating with different spatial frequencies and contrasts), texture-defined motion perception can be modeled in terms of a single motion energy channel.

Energy channels

Texture grabbers. Chubb and Sperling (1989a, 1991) suggested a two-stage mechanism for extracting texture-defined motion. Under their model, texture-defined motion is computed by motion energy channels whose initial transformations are called texture grabbers. As discussed below (see Rectification), a texture grabber is a linear spatial filter followed by rectification. In Stage 2, the time-varying output (activity) from each texture grabber is subjected to motion energy extraction.

Rectification. By rectification we mean any function that is zero for an input of zero, and is monotonically increasing for both positive and for negative real inputs.

Previously, Chubb and Sperling (1989b) demonstrated stimuli displaying systematic second-order motion that could be easily explained in terms of a texture grabber that used fullwave rectification (e.g. absolute value, square, etc.). However, the motion of these stimuli was inaccessible to any mechanism whose texture grabber used halfwave rectification (nonzero output only for positive or only for negative inputs). These results suggest that at least some of the texture grabbers used in second-order motion perception use fullwave rectification. It remains to be seen whether there are second-order motion mechanisms that use halfwave rectification. In the present context, however, we do not distinguish between different kinds of rectification. The essential nonlinear characteristic of texture extraction processes has also been recognized by Bergen and Adelson (1988) and Caelli (1985).

The linear filter used by a texture grabber is presumed to be realized in the visual system by an array of linear neurons, all with the same receptive field profile, distributed across the visual field. The texture grabber output results from applying some fixed, rectifying nonlinearity (e.g. the absolute value or the square) to the output of each of these linear neurons. It is assumed that the spatial filter of Stage 1 operates on stimulus contrast (see Model), rather than on luminance, but this assumption is not critical to our arguments. The output of a linear filter may be positive or negative depending on the local phase of the sensed texture. Thus the expectation of the output of such a filter is zero over the phase-randomized texture patches from which our stimuli are constructed. The purpose of rectification is to produce a positive average output across the texture so that a texture grabber registers the presence or absence of texture, independent of local phase. Indeed, that is why the Stage-1 transformation (linear spatiotemporal filter followed by rectification) is called a texture grabber.

Activity. The output of a texture grabber in response to a particular texture is called activity.

Motion energy-channels. Together, a texture grabber followed by motion energy extraction form one (texture-defined motion) energy channel.

Motion correspondence-channels. Together, a texture grabber followed by motion correspondence extraction form one (texture-defined motion) correspondence channel.

Previous research in texture-defined motion

Historically, motion correspondence has been investigated with ambiguous motion displays in which motion is perceived as occurring along one or the other of several competing paths. Most studies have dealt with stimuli that stimulated the first-order motion system (e.g. Burt & Sperling, 1981; Kolers, 1972; Navon, 1976; Papathomas, Gorea & Julesz, 1991; Shechter, Hochstein & Hillman, 1989; Ullman, 1980; Werkhoven, Snippe & Koenderink, 1990a; Werkhoven *et al.*, 1990b) and these data are adequately explained by the first-order motion energy extraction models.

We consider here two recent studies that attempt to deal with motion correspondence in texture-defined motion stimuli. These studies illustrate the difficult methodological issues that arise in attempting to determine motion correspondence, and thereby they indicate the necessity of the more complex paradigm which we use.

Watson's crossed-phi procedure. Watson (1986) attempted to measure the spatial frequency specificity of the perceptual mechanism responsible for texture-defined motion. He used a "crossed phi" method, in which two adjacent texture patches (A and B) in frame 1 exchanged positions in frame 2. The patches were Gaussian-windowed sine waves (Gabor patches). Observers reliably perceived apparent motion between the locations when A and B were different spatial frequencies. No apparent motion was reported when the patches were of similar spatial frequencies. Watson interpreted his results in terms of a model in which motion estimates are computed separately within different spatial frequency bands. He used the increasing probability of apparent motion with increasing differences in spatial frequency to estimate the spatial frequency selectivity of the motion channels. Furthermore, it was implicitly assumed that such a model was equivalent to a correspondence computation.

In our view, the ambiguous "crossed-phi" paradigm admits a simple alternative interpretation in terms of single energy channel model. Suppose there were just a single energy channel, and suppose that texture A happened to produce a bigger response from the texture grabber in this channel than texture B. Then, the change in position of patch A would produce a strong motion response in this channel; the change of position of patch B would produce a weak motion response in the opposite direction; net movement would be perceived in the A direction. The critical observation for a multichannel model is motion transparency—that motion of the A and B patches be seen simultaneously in opposite directions. Only then can we be sure that more than one channel is activated. In fact, such motion transparency was not reported by Watson, and, in our experience, it does not occur in such stimuli. Thus, Watson's experiment does not support a theory of multiple correspondence channels.

Green's Gabor patches. Green (1986) studied texture-defined motion with a rotating annular display similar to Navon (1976). The type of stimulus used by Green is schematized in Fig. 1. Call this stimulus *I*. One temporal period of *I* consists of four frames, as shown in Fig. 1.

Each of these frames is comprised of a circle of alternating patches of two types of texture, texture A and texture B. From frame to frame, these patches of texture take rotary steps clockwise around the circle. This rotary clockwise motion is equivalent to left-to-right motion in an analogous horizontal display, as indicated by the dotted lines connecting annular frames to horizontal frames.

Let T be an *arbitrary* texture grabber, and suppose that v_A is the average response of T to texture A and v_B is the average response of T to texture B. Then the output from texture grabber T in response to stimulus *I* is a spatiotemporal function whose average value over any patch containing texture A is v_A and whose average value over any patch containing texture B is v_B. Although there will certainly be variability to the T-output within a given texture patch, this intra-patch variability is not critical to the global motion percept elicited by *I*. What determines this global motion percept are the average T-output values, v_A and v_B, of patches of the two textures A and B.

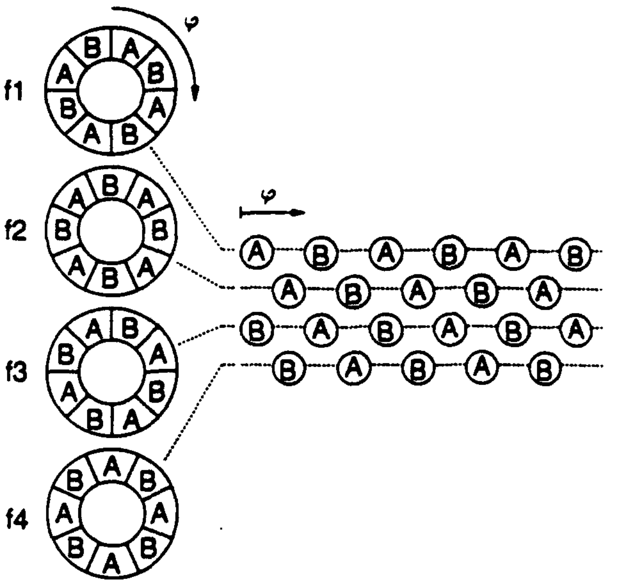


FIGURE 1. Green's stimulus, *I*. One temporal period of *I* consists of four frames. Each of these frames is comprised of a circle of alternating patches of two types of texture, texture A and texture B. From frame to frame, these patches of texture take rotary steps clockwise around the circle. This rotary clockwise motion is equivalent to left-to-right motion in an analogous horizontal display, as indicated by the dotted lines connecting annular frames to horizontal frames.

As many authors have observed (e.g. Adelson & Bergen, 1985; van Santen & Sperling, 1985), motion detection can be viewed as the detection of orientation in space-time. As is clear from inspection of Fig. 1(a), any motion detection mechanism that adheres to this general principle is bound to register clockwise motion in response to *I* whenever $v_A \neq v_B$.

In light of these observations, it is not surprising that observers in Green's experiment tended to perceive clockwise motion in displays such as *I*. In a critical sense, the clockwise motion of *I* is intrinsic to the format of the stimulus, and has little to do with the textures A and B comprising the patches of *I* (see Werkhoven *et al.*, 1990b). Nonetheless, Green took his results as support for the view that similar textures tend to match with each other in generating motion-from-texture.

Motion metamers

A psychophysical equivalence relation on a set Ω of physical stimuli is called a metamerism. Equivalent elements A and B of Ω are called metamers. Typically, metamerisms are defined using discrimination tasks. For example, if A and B are two illuminated patches that differ in spectral composition, we say they are metamers if an observer cannot distinguish between them.

In this paper, we focus on a different sort of metamerism that we call motion metamerism. Let Ω represent a set of texture patches that vary in spatial frequency, orientation, and contrast. The relation that we wish to capture is the following: for any two textures A and B in Ω, we call A and B motion metamers if and only if any occurrence of A in any dynamic visual display can be replaced by a patch of B without influencing the global motion percept elicited by that display. That is, A and B are motion metamers if and only if A and B are equivalent inputs to the mechanism that computes texture-defined motion. Obviously, A and B need not be equivalent inputs for other perceptual processes—as we shall show, motion metamers may appear quite different.

It is impractical to interchange A and B in all possible motion stimuli to verify that they are motion metamers. Instead, we use only two extreme test stimuli, in which any failure of metamerism would be most likely to appear. The essential core of the test we use is defined in terms of the stimuli I_1 and I_2 diagrammed in Fig. 2.

Each of these two stimuli pits two symmetrically opposite motion paths against each other. Stimulus I_1 pits a path comprised of a patch of texture A and a patch of texture B against a path comprised of two patches of texture A, whereas stimulus I_2 pits a path comprised of a patch of texture B and a patch of texture A against a path comprised of two patches of texture B. We presume that each of these paths has an associated motion strength, and that the global motion percept (left vs right) elicited by one of these stimuli depends only on which of its two paths has greater motion strength. In the case in which the global motion percept is ambiguous we assume that the strengths of the two component paths are equal.

For any textures A and B in Ω, we say A and B are transition invariant* if and only if the leftward vs rightward motion of each of I_1 and I_2 diagrammed in Fig. 2 is ambiguous (i.e. if each of I_1 and I_2 is equally likely to elicit a global rightward or leftward motion percept).

If textures A and B are transition invariant, then the motion strength of a match between A and A is equal to the motion strength of a match between A and B, and the motion strength of a match between B and A is equal to the motion strength of a match between B and B.

If A and B are motion metamers, then stimuli I_1 and I_2 are ambiguous in motion content; hence, A and B are transition invariant.

On the other hand, for practically all plausible texture-defined motion computations, if A and B are transition invariant, then they are also motion metamers. Indeed, the data we present make it clear that this is true of the computation that is actually used to compute texture-defined motion.

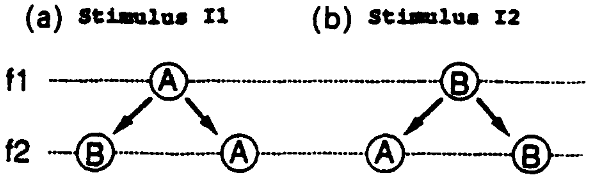


FIGURE 2. The binary relation ~ (transition invariance). (a) A schematic diagram of stimulus I_1. I_1 contains two frames. In the first frame there is a single patch of texture of type A. In the second frame, there are two patches of texture, one of type B and another of type A. These patches of texture are offset equal distances to the right and left of the location in frame 1 of the single patch of texture A. The stimulus I_1 sets up a competition between one motion path containing a patch of texture A and a patch of texture B and another, opposite motion path containing two patches of texture A. (b) A schematic diagram of stimulus I_2. For any textures A and B, we set A ~ B just if the stimuli I_1 and I_2 diagrammed in (a) and (b) respectively are both ambiguous in global motion content. That is, both stimuli I_1 and I_2 are equally likely to elicit global percepts of rightward or leftward motion. Any textures A and B for which A ~ B are said to be transition invariant. For a broad range of motion computations, it can be shown that, for any textures A and B, if A ~ B, then A and B are motion metamers in the strong sense (A and B can be freely traded for each other in any stimulus without changing the global motion percept elicited by that stimulus).

Motion competition schemes

The matching technique could be applied to a variety of ambiguous motion schemes for determining the dimensionality of the motion computation. However, not all of them have the power to discriminate between different types of motion channels (see e.g. the discussion on Green's display). We used an ambiguous motion scheme that was introduced by Werkhoven *et al.* (1990b). In this motion competition scheme, *one* heterogeneous motion path (between patches of texture s and texture v) competes *directly* with *one* homogeneous path (between patches of texture s).

By varying the properties of the textures v, we can determine the heterogeneous motion paths s, v that are equal in strength to a certain homogeneous path s, s.

Werkhoven *et al.*'s competition scheme not only allows to determine the dimensionality of the motion computation, but also allows to determine the number and type (energy vs correspondence) of channels involved in the motion computation. This requires a thorough analysis (given in the Model section).

However, an intuitively clear property of this scheme is that the two types of motion channels considered above (energy vs correspondence-channels) yield qualitatively different predictions for motion metamery and the relative strength of the heterogeneous and homogeneous motion paths. Hence, they are easily discriminated.

A preview

Dimensionality of the computation. In this paper, we discuss a general motion computation consisting of multiple motion channels, where each channel may be either an energy channel or a correspondence channel. By studying the above competition scheme with many

*The reason for this term will be clear in Transition Invariance and Motion Metamers.

different pairs of texture patches (Expts 1 and 2), we can determine classes of transition invariant textures (motion metamers) and infer the dimensionality of the motion computation (Model section). The results strongly support the view that texture-defined motion is computed by a single energy channel.

METHOD

In this section we describe the ambiguous motion competition scheme used in the experiments. This scheme (proposed by Werkhoven *et al.*, 1990b) differs from other schemes (e.g. Burt & Sperling, 1981; Green, 1986; Navon, 1976; Shechter *et al.*, 1989; Ullman, 1980) in that it contains a single heterogeneous motion path (between patches of texture 1 and texture 2) that competes directly with a single homogeneous motion path (between identical patches of texture 2). Except for textural properties, the other parameters (such as step size and frame rate) of the motion paths are identical.

Instead of varying both textures 1 and 2, we sampled a subspace of possible textures resulting in two (similar) schemes: Scheme I and Scheme II. In Scheme I, we kept texture 2 constant (now texture s) and varied texture 2 (now texture v).

Stimulus

Motion competition Scheme I. In Expt 1, we used motion competition Scheme I. The motion stimulus consisted of a series of eight frames ($f_1, f_2, \ldots, f_8$) shown successively in time. Figure 3 shows a sketch of the frames.

The first frame (f_1) contains an annulus of patches of alternating texture types s and v at regular positions (see Fig. 3, at the left side). Because the viewing distance was constant throughout the experiment, we will specify dimensions in degrees of visual angle. The annulus of texture patches has an inner radius of $r_1 = 1.04$ deg, and an outer radius of $r_2 = 2.08$ deg. The mean radius r is 1.56 deg. The patches (or *sectors*) are spatially contiguous. Since the annulus contains eight sectors, each sector has a width of 45 deg.

Frame f_2 was similar to frame f_1, except that patches of texture v are replaced by a uniform patch of background luminance. Furthermore, f_2 was rotated around the center of the annulus 22.5 deg with respect to frame 1 (see Fig. 3, left).

In a sequence of frames, the locations and types of patches in frame f_{n+2} were identical to frame f_n, except for a rotation around fixation of 45 deg.

The presentation time of a single frame ("frame-time") was 133.3 msec. Thus, the presentation time of the eight-frame sequence was 1.066 sec. The annulus revolved at an angular speed of 168.8 deg/sec, yielding a local velocity of the patch-centers of 4.6 deg of visual angle per second.

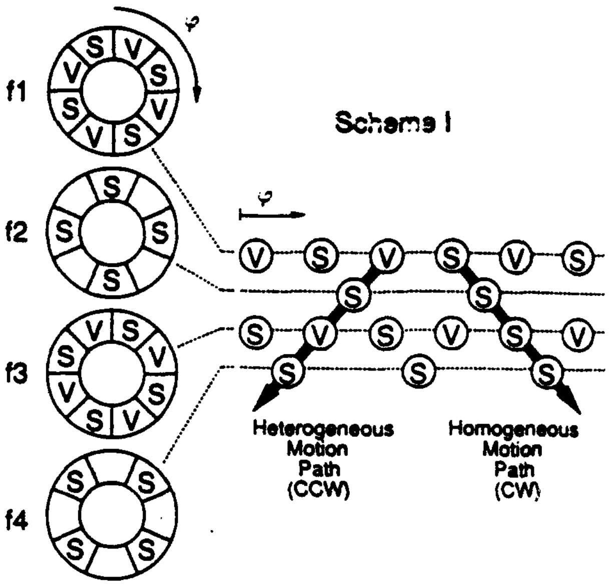


FIGURE 3. Motion competition Scheme I. Left: a series of frames ($f_1, f_2, \ldots$) is shown successively in time (for details see Method section). The first frame (f_1) contains an annulus of patches of alternated texture type s and v at regular positions drawn against a uniform background. The annulus has an inner radius of $r_1 = 1.04$ deg of visual angle, and an outer radius of $r_2 = 2.08$ deg. The patches of texture s and texture v are spatially contiguous and alternate within the annulus. Since the annulus contains eight patches, each patch has a width of 45 deg. Angular position φ is measured clockwise with respect to the vertical. The second frame (f_2) is similar to frame f_1, except that the low frequent patches of texture v are now replaced by a uniform patch of background luminance. Furthermore, f_2 is rotated (clockwise) around the center of the annulus over an angle of 22.5 deg with respect to frame f_1. In a sequence of frames, frame f_{n+2} is identical to frame f_n, except for a rotation around the center over an angle of 45 deg (clockwise). Right: angular positions φ is along the horizontal axis. Patches of texture s and v are shown at their angular positions for frames $f_1 \cdots f_4$ yielding rows of patches. The top row of patches s and v corresponds to frame f_1. The second row of patches s corresponds to frame f_2. Hence, time (or frame number) is along the vertical axis. When frame f_n and frame f_{n+1} are presented in succession, two motion paths are *a priori* likely. A homogeneous motion path: clockwise matches (CW) between patches of identical texture s (indicated by the arrow pointing down and right). A heterogeneous motion path: counter-clockwise (CCW) matches between patches of texture s and patches of texture v (indicated by the arrow pointing down and left).

The ambiguous motion stimulus described above contains two motion paths. This can be understood most easily using a diagram in which we show the angular positions (φ) of the patches of texture for successive frames. Angular position is measured clockwise relative to the vertical. Such a diagram is shown in Fig. 3, at the right side. Note that the horizontal rows of patches correspond to frames 1, 2, 3 and 4 respectively. By definition, motion extraction is based on the dynamic properties of the stimulus, that is the spatiotemporal pattern of textures. In the diagram, possible motion paths are spatiotemporal (oblique) rows of elements. The arrows pointing to the left and right are examples of motion paths to the left and right respectively. In the following descri[illegible] f the stimulus, we will say that the neighboring ele[illegible] n a motion path are spatiotemporally linked [illegible] matched". Note that the term "matching" is used for the purpose of stimulus description only and that it does *not* refer to a "motion correspondence" computation.

When frame f_n and frame f_{n+1} were presented in succession, two matches between patches of frame f_n and patches of frame f_{n+1} were *a priori* possible. The first match is a homogeneous clockwise match between patches of identical texture s separated by +22.5 deg (indicated in the diagram by the arrow pointing down and to the right). The second match is a heterogeneous counter-clockwise match between patches of texture v and patches of texture s (−22.5 deg, indicated by the arrow pointing down and to the left). Matches between frames f_n and f_{n+2} are entirely ambiguous. Matches between patches of frames f_n and f_{n+3} involve large temporal separations (400 msec) relative to the equivalent matches between frames f_n and f_{n+1} (133.3 msec). It has been shown that motion strength decreases strongly and monotonically with temporal interval for intervals larger than approx. 30 msec (Burt & Sperling, 1981; Werkhoven & Koenderink, 1991). Therefore, the matches between frames f_n and f_{n+3} are unimportant for motion perception in these stimuli.

Scheme I displays contain homogeneous and heterogeneous motion paths in opposite directions. By randomizing the direction of rotation, the directions of the two motion paths (although still opposite) are randomized.

The annular pinwheel stimulus was used for various reasons. First, the motion stimulus was presented at a constant eccentricity in the parafovea, and the effects of anisotropy of the retina were averaged across equivalent areas of the visual field. Second, it was easier to maintain fixation so eye movements were better controlled.* Finally (with the use of circularly symmetric stimuli) a motion path does not end at the boundaries of the display, avoiding edge effects.

Motion competition Scheme II. Scheme II (used in Expt 2) is equivalent to Scheme I, except that textures s and v are interchanged. The motion stimulus and the resulting motion paths for this experiment are shown in Fig. 4.

Although the heterogeneous motion path (between patches of texture s and v) is identical to that of Scheme I, the homogeneous motion path is different from that of Scheme I. In Scheme II, the homogeneous motion path consists of patches of texture v. The critical importance of the two schemes for our paradigm concerns the question of whether, when a particular s and v are chosen so that motion paths are balanced in Scheme I, the paths will remain balanced when the same s and v are used in Scheme II. From the subjects' point of view, however, there is no difference between the two schemes because, for any stimulus generated by Scheme I, an identical stimulus can be generated by Scheme II.

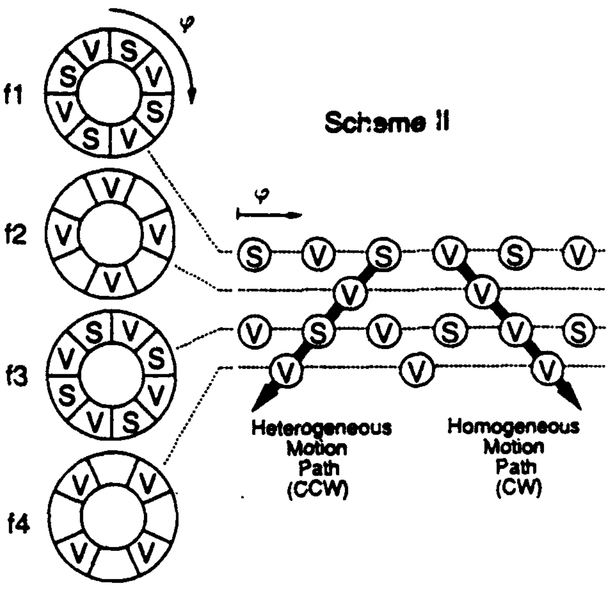


FIGURE 4. Motion competition Scheme II. This scheme is similar to Scheme I (see Fig. 3), except that textures s and v are interchanged. In Scheme II, the homogeneous motion path contains textures v.

However, during the course of a session, when v is varied between trials, different families of stimuli are generated by the two schemes.

Texture stimuli

The textures used to characterize texture-defined motion are patches of sinusoidally modulated gratings that differ in spatial frequency and amplitude. The grating patches were arranged in eight sectors of an annulus (pinwheel) around the fixation point with the grating extending radially in each sector. Two critical parameters that characterize a texture patch at a given location of the pinwheel are amplitude m and spatial frequency ω. Within a location, grating orientation was always radial. The phase γ of the grating was a random variable with a uniform distribution.

We use polar coordinates to further characterize the pinwheel. Let φ be the polar angle of a point in the image, and ρ be the distance to the origin (the center of the annulus). Then the luminance distribution at the point ρ, φ in sector j of frame i is:

$$L_{i,j}(\rho,\varphi) = L_0[1 + m_{i,j}\sin(2\pi r\varphi\omega_{i,j} + \gamma_{i,j})]. \quad (1)$$

We define the mean spatial frequency $\omega_{i,j}$ as the spatial frequency at mean radius r. The mean spatial frequency $\omega_{i,j}$ of a texture patch depends only on whether j is odd or even. That is, two spatial frequencies, ω_s, ω_v strictly alternate between adjacent patches on every frame of the display.

Within a trial, the amplitude $m_{i,j}$ of a sector i,j depended only on whether i and j were even or odd. On odd frames, $m_{0,j}$ was chosen as m_s or m_v according to whether the sector j was even or odd. On even frames, sector amplitude $m_{e,j}$ alternated between 0 and m_s in Scheme I and between m_v and 0 in Scheme II. Between trials, m_v and ω_v were changed. Sixteen values of

*Torsional eye-movements induced by the rotating annuli (cycloinduction) were not controlled in our experiment. Balliet and Nakayama (1978) reported the ability of extremely trained subjects to make stepwise eye torsions up to rotations of approx. 26 deg for large field stimuli (25–50 deg of visual angle). However, we do not expect torsional pursuit in our experimental conditions: small field stimuli, brief presentations, fast motion, unpredictable motion direction, and ambiguous or near-threshold motion stimuli.

amplitude m_v from 0 to 1 were used increasing by steps of 0.0625: 0, 0.0625, 0.13, ..., 1. Spatial frequency ω_v was varied over a range of three octaves: 1.2, 2.5, 3.7, 4.3, 4.9, 5.6, 7.4 or 9.9 c/deg. The amplitude m_s and spatial frequency ω_s of texture s were constant throughout the experiment: $m_s = 0.5$, $\omega_s = 4.9$ c/deg.

The phase $\gamma_{i,j}$, $0 \leqslant \gamma_{i,j} \leqslant 2\pi$, was chosen randomly and independently for every combination of i and j, that is, for every single patch. The phase randomization of every patch makes the motion of the stimulus inaccessible to any first-order (Fourier-based) mechanism. Phase randomization insures that motion mechanisms sensitive to correspondences in stimulus luminance were not systematically engaged (Chubb & Sperling, 1988).

Figure 5 shows an example of a series of frames for Scheme I. Texture s is a "medium" frequency grating and texture v is a "low" frequency grating. The regions inside and outside the annulus (background) were uniform gray and had a luminance value ($L_0 = 72$ cd/m^2). Within the annulus' texture patches the expected luminance value was equal to the background luminance.

Apparatus

The experiment was controlled by a IBM 386 PC compatible computer, driving a TrueVision AT-Vista video graphics adapter. A 60 Hz Imtec 1261L monitor with a P4-type phosphor was used to display the stimuli. The screen dimensions were 21.8 × 14 cm (640 × 480 pixels; 12.3 × 8.0 deg visual angle).* We used a look-up table to linearize the monitor's luminance values with the gray values of the computed stimulus patterns. The decay time to 10% and 1% intensity was about 1.3 and 6.2 msec respectively which is shorter than the temporal properties of retinal processing (Farrell, Pavel & Sperling, 1990; Sperling, 1976).

Subjects

Two subjects participated in the experiments: one of the authors (PW) and a colleague (JS). PW is emmetropic. JS is myopic (−0.5 D) but was in focus for the viewing distance used. Both subjects were experienced psychophysical observers. Natural pupils, binocular viewing, and spectacle corrections were used throughout. Several naive subjects confirmed the main findings for the experiments.

Procedure

Subjects indicated the dominant motion path (counter-clockwise/clockwise) by pressing one of two buttons. In both experiments, texture s (the standard texture) had amplitude $m_s = 0.5$ and spatial frequency $\omega_s = 4.9$ c/deg. From trial-to-trial, the spatial frequency ω_v and amplitude m_v of texture v was varied. The experiments determined the probability $P_i(m_v; \omega_v)$ of perceptual dominance of the heterogeneous motion path as a function of m_v for certain ω_v using the method of constant stimuli. The subscript i, $i = 1, 2$, indicates Expt 1 with competition Scheme I (Fig. 3) or Expt 2 with Scheme II (Fig. 4).

The probabilities $P_1(m_v; \omega_v)$ and $P_2(m_v; \omega_v)$ are estimated by the fraction of perceptually dominant heterogeneous motion paths out of 36 presentations. Spatial frequency ω_v was varied over a range of three octaves: $\omega_v = 1.2, 2.5, 3.7, 4.3, 4.9, 5.6, 7.4$ and 9.9 c/deg. Within a session, amplitude m_v was varied (pseudo-randomly from trial-to-trial; ω_v was varied only between sessions. For each spatial frequency ω_v, Expts 1 and 2 were both conducted within one session.

Subjects viewed the stimuli in a room with dimmed background illumination.

EXPERIMENT 1: SCHEME I

Results

By definition, the homogeneous path (consisting entirely of identical patches of texture s) does not change in this experiment when texture v is varied (see Scheme I, Fig. 3). The strength of the heterogeneous path, which is composed of alternate patches of textures s and v is varied by varying spatial frequency and amplitude, ω_v and m_v, of texture v. Figure 6 shows the probability $P_1(m_v; \omega_v)$ of reporting the heterogeneous motion path as dominant as a function of the amplitude m_v of texture v. Each panel shows $P_1(m_v; \omega_v)$ for a different value of spatial frequency ω_v.

The data show that the probability of reporting the heterogeneous path as dominant increases monotonically from 0 (for small m_v) to 1 (for $m_v = 1$) for all values of ω_v except the highest, where the probability of heterogeneous motion dominance has only reached about 65% when $m_v = 1$. A remarkable feature of these data is that in all eight panels, the probability $P_1(m_v; \omega_v)$ of heterogeneous motion dominance exceeds 50% for sufficiently high amplitude of patch v.

The upper left panel of Fig. 6 shows data for a two octave difference between the spatial frequency of texture s ($\omega_s = 4.9$ c/deg) and the spatial frequency of texture v ($\omega_v = 1.2$ c/deg). Heterogeneous motion is perceived in 50% of the presentations when the amplitude m_v of texture v is approx. 0.2. Note that at this balance point where both paths are equally likely, both the amplitudes and the spatial frequencies of textures s and v are markedly different. Once m_v exceeds 0.5, the heterogeneous motion path is dominant in 100% of the presentations. A 100% perceptual dominance of a heterogeneous over a homogeneous path demonstrates that the similarity between the textures in a motion path certainly is not essential for motion strength. Indeed, for sufficiently large m_v, the heterogeneous path is dominant over the homogeneous path for every combination of frequencies tested in Fig. 6.

*Due to the limited bandwidth of the video amplifier (30 MHz) of the monitor, an anisotropy was observed for the average luminance of differently oriented textures that contain high spatial frequencies. Therefore, we only displayed the pixels at column position m and row position n for which $(m + n)$ was even. The other pixels were dark. Hence, vertical and horizontal gratings share a common "carrier' component. This procedure forfeits maximum luminance and resolution in favor of eliminating anisotropy; the net resolution (320 × 240 pixels) was more than adequate for the displays.

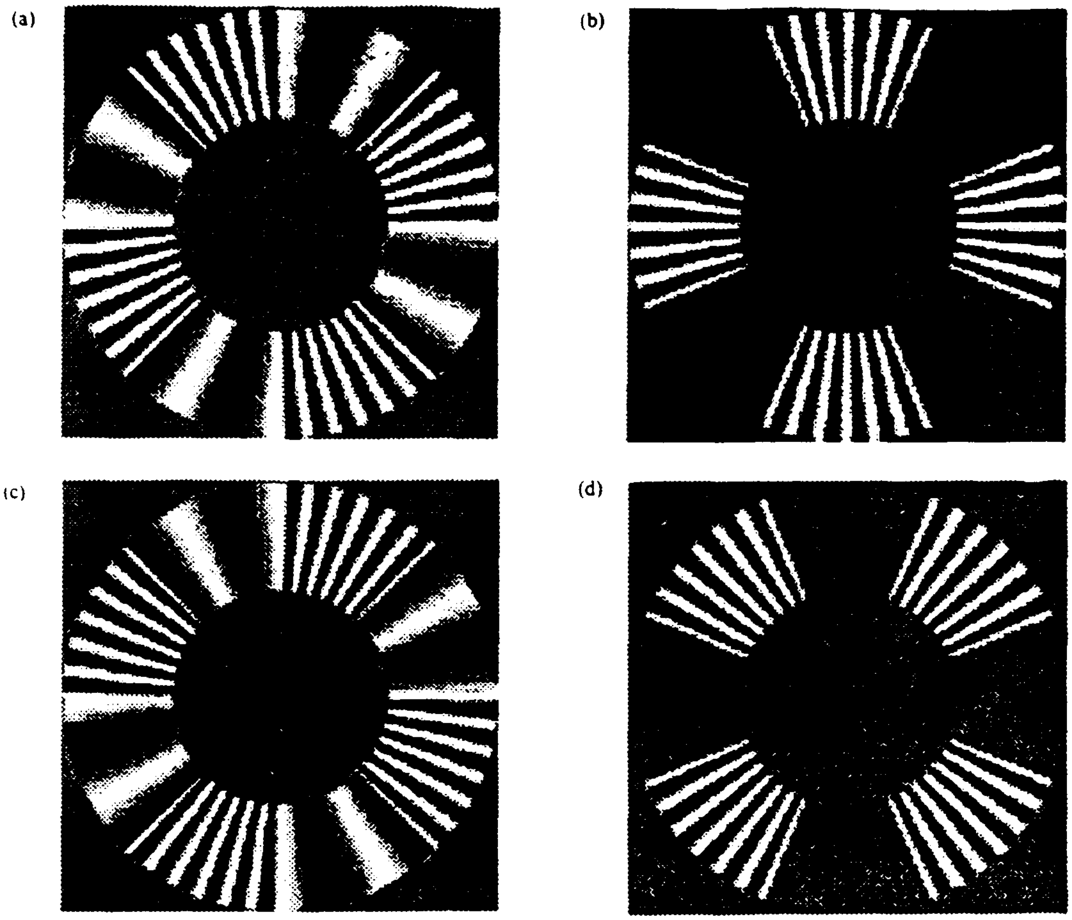


FIGURE 5. An example of the ambiguous motion display (as sketched in Fig. 3). Frames f_1, f_2, f_3, and f_4 (containing the patches of textures) are shown in (a), (b), (c) and (d) respectively. For this example, textures s and v differ only in their spatial frequency: the spatial frequency of texture s is two octaves higher than that of texture v.

The transition amplitudes between heterogeneous and homogeneous motion occur where the curves of Fig. 6 cross 50%. The transition amplitudes occur at a wide range of different amplitudes m_v for different spatial frequencies ω_v. Each P_1 curve is well characterized by two parameters: the transition amplitude $\mu_1(\omega_v)$ and the steepness $\sigma_1(\omega_v)$ at the transition amplitude (the subscript 1 indicates Scheme I). The transition amplitude $\mu_1(\omega_v)$ is defined as the amplitude m_v of texture v, necessary for balancing the motion paths [such that $P_1(m_v; \omega_v) = 50\%$], the steepness $\sigma_1(\omega_v)$ is defined as the derivative $\partial/(\partial m_v)P_1(m_v; \omega_v)$ with respect to m_v at the transition amplitude.

To estimate transition amplitude $\mu_1(\omega_v)$ and steepness $\sigma_1(\omega_v)$, we selected* data points of each probability curve around the transition amplitude. Within this selected range, the curve was assumed to be linear, and these data points were subject to a least square method of linear regression to estimate the regression coefficients $\mu_1(\omega_v)$ and $\sigma_1(\omega_v)$.

Estimates of $\mu_1(\omega_v)$ are shown in Fig. 7 as a function of the varied spatial frequency ω_v (open circles). The transition amplitude $\mu_1(\omega_v)$ increases systematically with increasing spatial frequency ω_v of texture v for both subjects. Together, the data of Figs 4 and 5 indicate that the strength of the heterogeneous motion path increases with increasing amplitude m_v but decreases with increasing spatial frequency ω_v.

Estimates of $\sigma_1(\omega_v)$ are shown in Fig. 8 as a function of the varied spatial frequency ω_v (open circles). The steepness $\sigma_1(\omega_v)$ of the probability curves at transition amplitude $\mu_1(\omega_v)$ decreases with the spatial frequency ω_v of texture v. In the Model section we elaborate on this finding.

Discussion

Sufficiency of a single energy-channel. In a single energy-channel, we assume that only one single type of texture grabber operates on the input yielding an activity representation of the input. Motion strength is the result

*In principle, we selected the three data-points around the transition amplitude (the crossing of the curves with the 50% guide line) that were closest to the 50% guide line. There were only two exceptions. First, at spatial frequency $\omega_v = 1.2$ c/deg, for subject PW, Expt 2, we selected the data points with amplitude $m_v = 0.19$, 0.25 and 0.31 (to avoid the low amplitude values, for which Scheme II becomes ambiguous). Second, at spatial frequency $\omega_v = 2.5$ c/deg, for subject JS, Expts 1 and 2, we selected the data points with amplitude $m_v = 0.38$ and 0.5 (since we had no data points close to the guide line).

of a motion energy analysis scheme applied to this activity representation. The motion strength of a path is computed from the product of activity measures between successive patches along the path in space-time. Motion strength of a heterogeneous path balances homogeneous motion strength when the responses (activities) to textures v and s are equal. Differences in textural properties between elements s and v are irrelevant as long as the activities are equal, just as, in scotopic vision, differences in wavelength are irrelevant as long as the rod response is the same.

The results for Scheme I suggest an activity transformation that is a monotonically increasing function of amplitude and a monotonically decreasing function of spatial frequency. For example, to balance the activity of texture s, with amplitude m_s and spatial frequency ω_s, with a lower spatial frequency texture v, $(m_v; \omega_v)$ requires a $m_v < m_s$. This pattern of results suggests a single class of texture grabbers consisting of a low-pass spatial filter followed by rectification.

We argued that a single energy-channel is sufficient to explain the results of Expt 1. It is important to note here, however, that our finding that heterogeneous motion can dominate homogeneous motion is also consistent with multiple energy-channels, as will be shown in the Model section. For example, the dominance of heterogeneous motion may well be the result of two independent energy-channels, both favoring heterogeneous motion. To uniquely determine the number of channels involved, we need the results for competition Scheme II together with a formal analysis (Model section).

Secondary contributions of a correspondence-channel. In the Discussion above, we argued that a single-channel model is sufficient to model the (amplitude/frequency dependent) dominance of heterogeneous motion found for Scheme I. However, we cannot exclude a possible secondary effect of texture similarity based on this scheme. To motivate Expt 2, we need to elaborate on this argument.

Although motion perception may be dominated by a single energy-channel, there may yet be a secondary contribution of a correspondence-channel.

The relative strength of the heterogeneous motion path would decrease as the differences between the spatial frequencies and amplitudes of successive patches of textures s and v increased. Suppose there were a

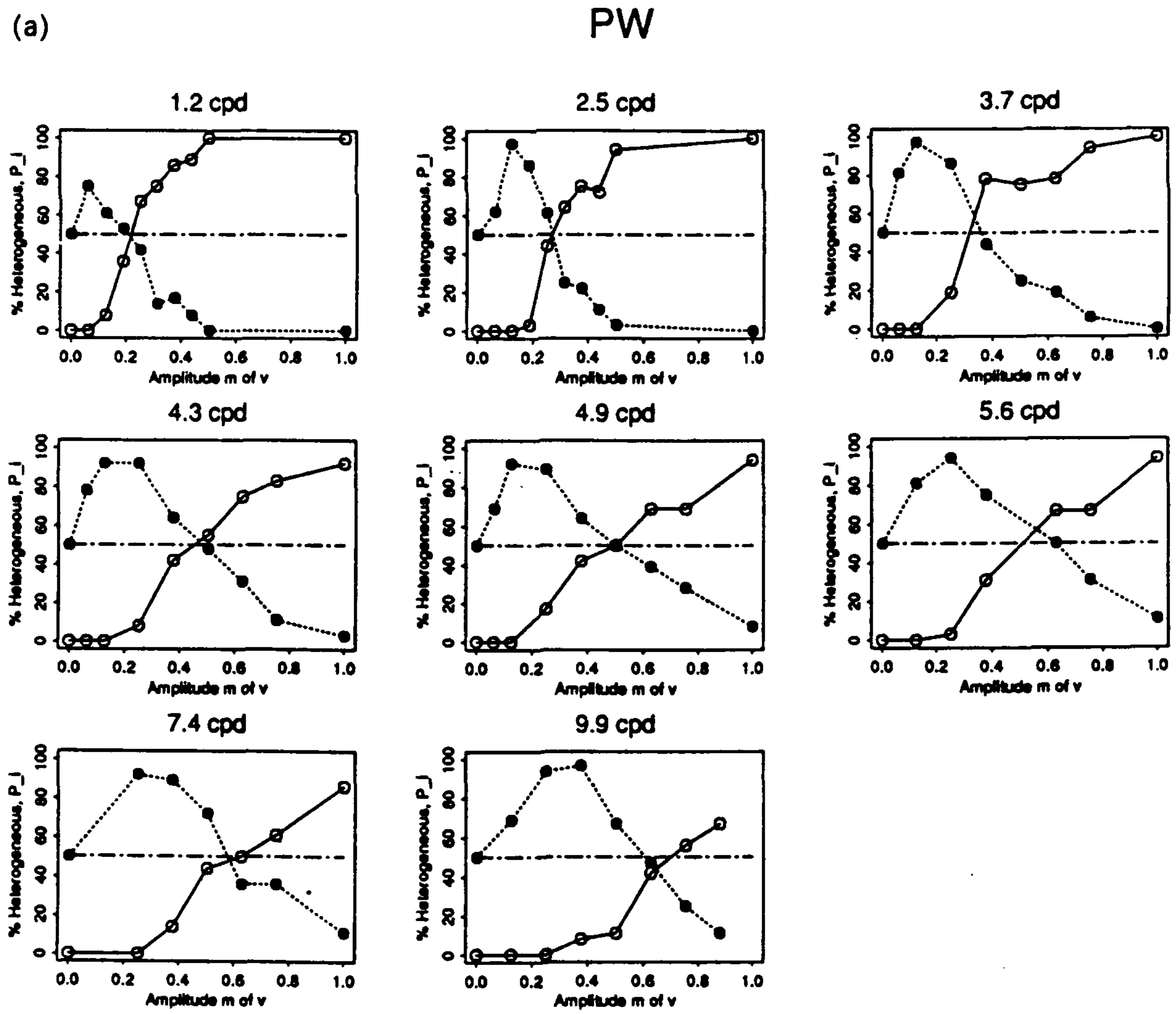


FIGURE 6(a). *Caption overleaf.*

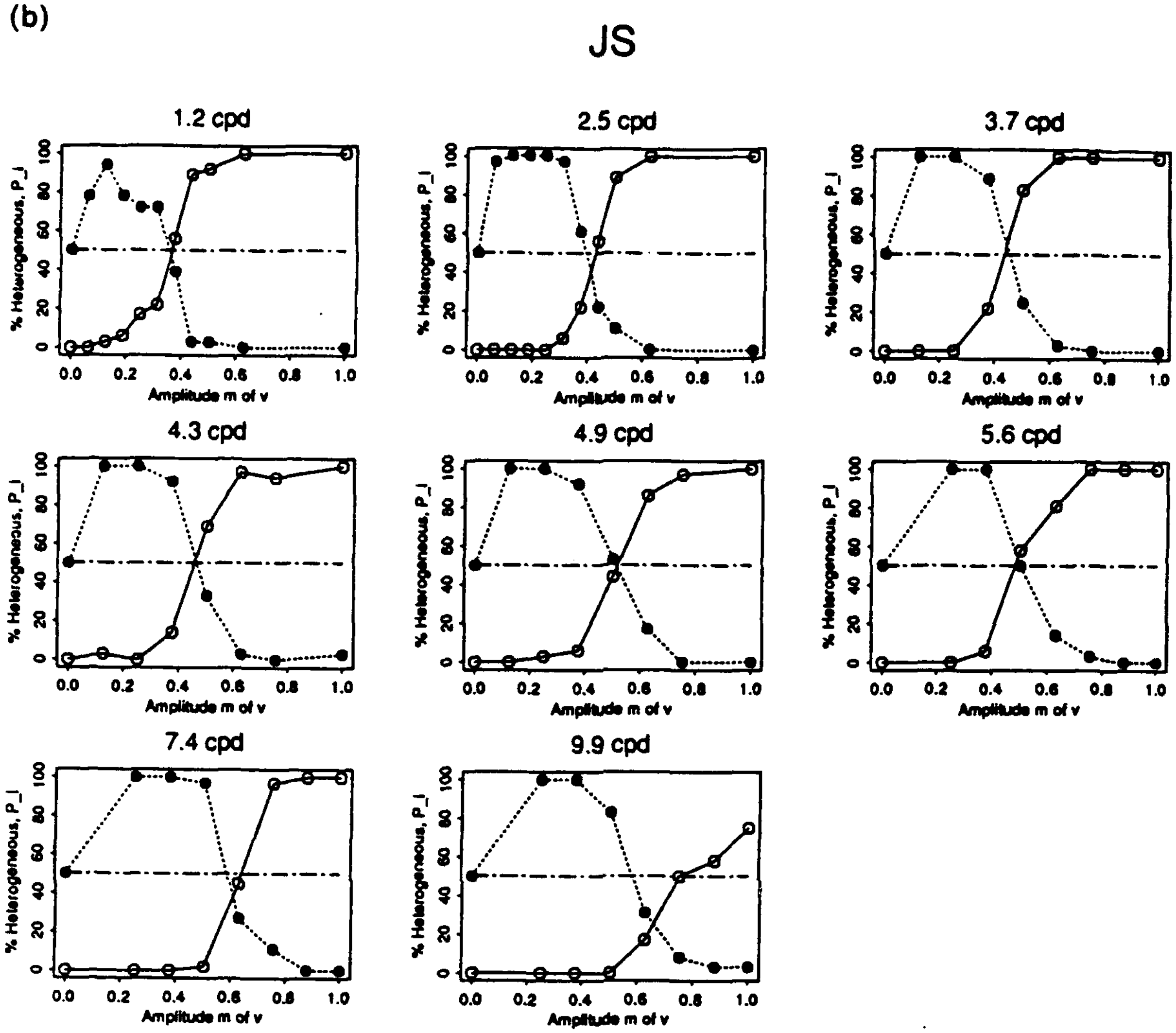


FIGURE 6. Probability $P_i(m_v; \omega_v)$ of dominance of a heterogeneous motion path over a homogeneous motion path is shown as a function of the amplitude m_v of texture v for different spatial frequencies ω_v of texture v for two subjects. Open circles represent the probability $P_1(m_v; \omega_v)$ for Scheme I (Fig. 3); solid circles $P_2(m_v; \omega_v)$ for Scheme II (Fig. 4). The horizontal dashed guide line indicates a 50% probability of heterogeneous motion dominance. The amplitude m_s and spatial frequency ω_s of texture s is the same for all panels: $m_s = 0.5$ and $\omega_s = 4.9$ c/deg. (a) Subject PW; (b) subject JS.

secondary contribution of a correspondence-channel in Expt 1, sensitive to differences between textures in either amplitude or frequency. Because the correspondence-channel favors the homogeneous path (by definition), motion balance requires v in the heterogeneous path to have a higher amplitude m_v to overcome the similarity in path s, s than if there were no correspondence-channel. Thus, in Scheme I, a secondary correspondence effect would displace transition amplitude $\mu_1(\omega_v)$ to higher values.

To test for a correspondence-channel, we introduce Scheme II in which s and v are interchanged (see Fig. 4). If there were a correspondence effect, in Scheme II it would favor the v, v path and the transition amplitude $\mu_1(\omega_v)$ would be shifted below $\mu_2(\omega_v)$ for any texture v.

When the homogeneous and heterogeneous motion paths remain balanced after interchanging textures s and v, this is called transition invariance. Transition invariance would imply that there is no contribution of a correspondence-channel.

EXPERIMENT 2: SCHEME II

Results

Figure 6 shows the probabilities $P_2(m_v; \omega_v)$ of the dominance of the heterogeneous motion path as a function of the amplitude m_v of texture v for different spatial frequencies ω_v of texture v. The data points for Scheme II are marked by a solid circle.

When $m_v = 0$, the display is physically as well as perceptually ambiguous. A value of 50% is shown for $m_v = 0$, though no data were collected at this point. By varying the amplitude of texture v in this experiment, the strength of both the heterogeneous motion path and the homogeneous motion path are varied. As the amplitude m_v increases, the probability of heterogeneous motion

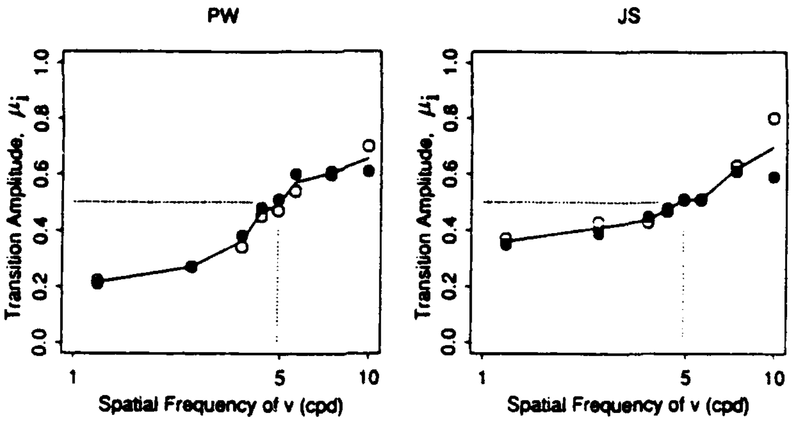


FIGURE 7. Transition amplitudes $\mu_i(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I, solid circles for Scheme II. The vertical dashed line indicates the spatial frequency of texture s: $\omega_s = 4.9$ c/deg. The horizontal dashed guide line indicates the amplitude of texture s: $m_s = 0.5$.

dominance first increases to a maximum, then decreases to zero for high amplitude m_v. On the whole, for amplitudes above 0.1 or, in a few cases, 0.2, the Scheme I and Scheme II curves are mirror complementary, and seem to cross at exactly $P = 50\%$. That is, the two schemes produce remarkably similar transition amplitudes.

To examine the correspondence between the data from Schemes I and II, some definitions are needed. Let the transition amplitude $\mu_2(\omega_v)$ be the amplitude m_v of texture v for which the motion paths are balanced, and the probability of heterogeneous motion dominance $P_2(m_v; \omega_v)$ is 50%. The steepness at this transition amplitude is $\sigma_2(\omega_v)$. The transition amplitude $\mu_2(\omega_v)$ and steepness value $\sigma_2(\omega_v)$ are estimated as $\mu_1(\omega_v)$ and $\sigma_1(\omega_v)$ in the previous section.

To compare the transition amplitude $\mu_2(\omega_v)$ for Scheme II with transition amplitude $\mu_1(\omega_v)$ for Scheme I, they are presented together as a function of spatial frequency ω_v in Fig. 7. Transitions $\mu_2(\omega_v)$ are presented with solid circles. As in Scheme I, the amplitude $\mu_2(\omega_v)$ of texture v, necessary for balancing the motion paths, increases systematically with increasing spatial frequency ω_v of texture v. An exception for both subjects are the transition amplitudes for $\omega_v = 9.9$ c/deg.

To compare the steepness values $\sigma_2(\omega_v)$ for Scheme II with steepness values $\sigma_1(\omega_v)$ (for Scheme I), the absolute value of $\sigma_2(\omega_v)$ is shown as a function of the varied spatial frequency ω_v in Fig. 8 (using solid circles). It should be noted that the estimation is not very accurate: the standard deviation in the distribution of steepness coefficient $\sigma_i(\omega_v)$ is approx. 20%. However, like $\sigma_1(\omega_v)$, the steepness $\sigma_2(\omega_v)$ shows a tendency to decrease with increasing spatial frequency ω_v of texture v.

Discussion

Transition invariance and motion metamers. It is immediately clear that, for most spatial frequencies ω_v of texture v, the transition amplitude $\mu_2(\omega_v)$ is equal within measurement error to transition amplitude $\mu_1(\omega_v)$ (see Fig. 7). In fourteen of sixteen cases, the transition amplitudes are invariant when the textures s and v are interchanged. This we call transition invariance.

In two cases (the highest spatial frequency used—$\omega_v = 9.9$ c/deg—for both subjects), a small difference between transition amplitudes for Schemes I and II is observed. At the high spatial frequency of v, the amplitude of texture v necessary to balance the motion paths is slightly smaller for Scheme II than for Scheme I. This shift in transition amplitude suggests a small

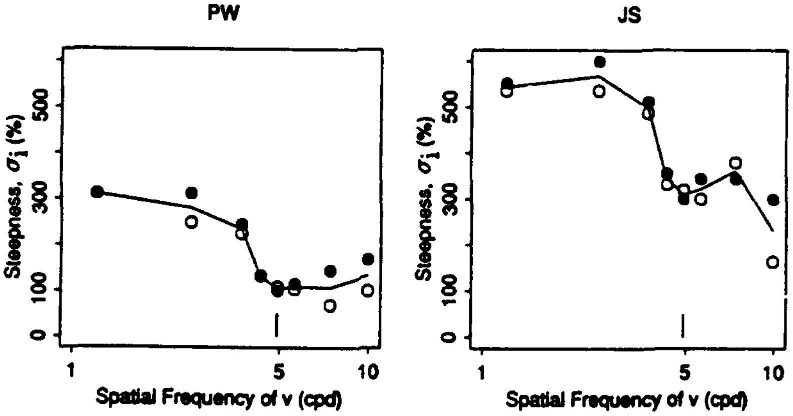


FIGURE 8. Steepness values $\sigma_i(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I, solid circles for Scheme II. (Note that to facilitate comparison absolute values are given!) The vertical dashed guide line indicates the spatial frequency of texture s: $\omega_s = 4.9$ c/deg.

similarity effect (a small contribution of a correspondence-channel), and was discussed in the Discussion of Expt 1.

Transition invariance implies that textures s and v (at transitions) are equivalent with respect to motion processing and can be interchanged in any motion path (Scheme I and Scheme II) without affecting motion strength. This leads to the important conclusion that textures s and v are (texture-defined) motion metamers.

It is interesting to note that Green (1986, Fig. 7, p. 604) was unable to find an amplitude that could make a spatial frequency patch of 5.0 c/deg into a motion metamer of a 1.7 c/deg patch. We had no difficulty in finding metamers between even more disparate spatial frequencies. However, our data in Fig. 5 show that one of the two subjects would require the 5 c/deg stimulus to have more than two times the amplitude of the 1.7 c/deg stimulus, and this is outside the range of amplitudes that Green explored.

Necessity of a single energy-channel. The general finding of transition invariance strongly constrains the possible ways in which motion can be computed between textures in the class we are considering.

Transition invariance shows that there is *no* secondary contribution of correspondence-channels (see the discussion on this issue in Expt 1). The effect that a patch of texture v has on the strength of motion is independent of the other patches in the path. At a transition, the strength of motion path s, v is equal to that of v, v and that of s, s, although a correspondence-channel would yield stronger motion for the homogeneous paths.

The only alternative is a system of multiple energy-channels that must be combined and represented by a single scalar representation (e.g. summation of energy-channels). In the Model section, we prove (under the assumption of channel summation) that if multiple energy-channels were involved, the transition amplitude would generally shift when the textures s and v are interchanged in Schemes I and II. However, when motion perception is exclusively ruled by a single energy-channel (the product of the activity of a single type of texture grabber), the transition amplitude is invariant when the textures s and v are interchanged. Hence, transition invariance uniquely supports a single energy-channel model of texture-defined motion perception.

EXPERIMENT 3: AMPLITUDE LINEARITY

Motivation

In the above experiments, we have shown that the transition amplitude $\mu_t(\omega_v)$ increases systematically with increasing spatial frequency ω_v of texture v for both subjects. The strength of the heterogeneous motion path in Scheme I increases monotonically with increasing amplitude m_v but decreases with increasing spatial frequency ω_v. In order to further specify the dependency of motion strength on amplitude, we performed an experiment similar to that described above using competition Scheme I, and varied the amplitude of texture s.

Results

We kept the frequency of textures s and v constant ($\omega_s = 4.8$ c/deg and $\omega_v = 1.2$ c/deg) and measured the transition amplitude μ_t as a function of amplitude m_s (Scheme I). Transition amplitude was estimated from the psychometric curves using the method described earlier.

Figure 9 shows the transition amplitude μ of texture v for three amplitude values of texture s ($m_s = 0.50, 0.75$ and 1.00) for three subjects. The data strongly suggest a linear dependence of the transition amplitude of texture v on the amplitude of texture s. The solid lines are the best fits (minimizing the sum of squares), accounting for at least 97% of the variance for each subject.

Discussion

We showed that the transition amplitude of texture v needed to balance the motion path s, v with the motion path s, s varied linearly with the amplitude of texture s. This dependency is easily accommodated in a model where the texture grabber is linear in the amplitude of the texture. In fact, one can easily show that amplitude linearity follows directly from the linear data under the assumption that the texture grabber is a separable function of spatial frequency and amplitude. A linear (low-pass) spatial frequency filter is a simple example of such a separable filter characteristic.

MODEL

Summary of model constraints

We used the analogy with colorimetry and some general assumptions about the possible motion computations involved to reach the conclusion that texture-defined motion strength is ruled by a single energy-channel. We summarize our reasoning.

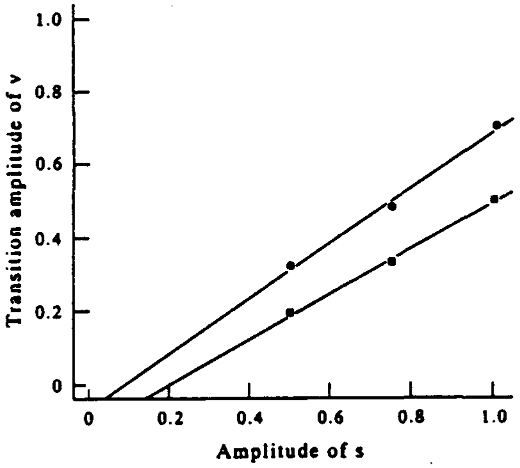


FIGURE 9. The dependence of transition amplitude μ_t (ω_v) on amplitude m_s of texture s. The spatial frequency ω_s was 4.9 c/deg, and ω_v was 1.2 c/deg. Competition Scheme I was used. Circles, subject JS; squares, subject PW. The solid lines show the best linear fit (minimizing the sum of the squared deviations).

We discriminate two classes of motion computations: energy-channels and correspondence-channels, yielding different metrics for the strength of a motion path. Consider, a heterogeneous motion path composed of patches of texture s and v. The strength of an energy-channel for an s, v path is determined by the product of the activity of texture s and that of v. The activity of a texture is the output of some nonlinear transformation (texture grabber) that maps texture into a scalar. Energy-channels are insensitive to differences in textural properties and allow heterogeneous motion paths s, v to dominate over homogeneous paths. By definition, the strength of a correspondence-channel is determined by the similarity of the textural properties of textures s and v. That is, homogeneous paths s, s and v, v dominate heterogeneous paths s, v.

In theory, multiple channels of each type may be involved in a motion computation yielding a motion strength vector representation of arbitrary dimensionality. However, the experimental results impose the following constraints. First, the class of motion paths equal in strength for both Scheme I and Scheme II indicates that the computation is one-dimensional. Second, the invariance of transitions for Scheme I and Scheme II exclude correspondence-channels. This leaves us with a system of multiple energy-channels, that combine into a single scalar representation of motion strength.

Although we have shown that a single energy-channel is sufficient to model the data, we promised a proof for the necessity of a single energy-channel. This proof is based on the inconsistency of multiple energy-channels with transition invariance. We assume a system of multiple energy-channels that linearly combine to represent motion strength (summation of energy-channels). Such a system would result in different transitions for Scheme I and II. The proof is given and discussed in the Appendix.

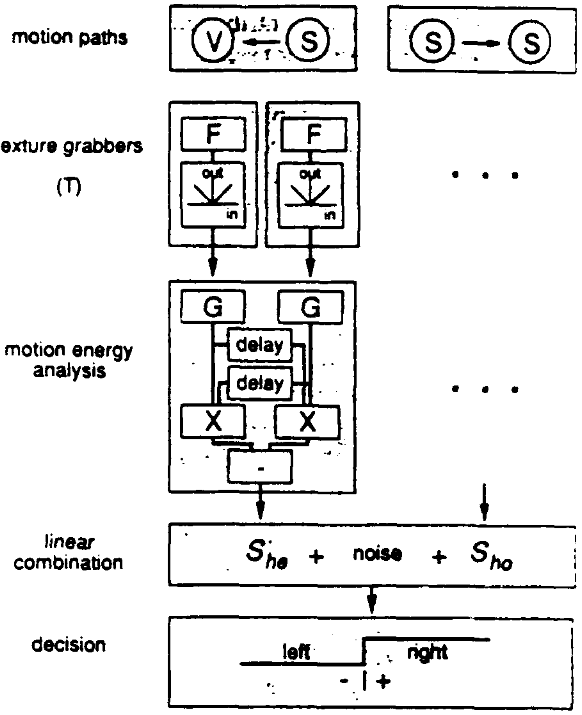


FIGURE 10. Diagram of a single channel motion computation. First stimulus amplitude is extracted followed by a linear spatial filter F and rectification. The spatial filter together with the rectification is called "texture grabber" (the first stage). The output of the texture grabber is called activity. The second stage (motion energy analysis) is basically a coincidence detector: it computes the product of the delayed activity at location 1 with the current activity at location 2. Response variability across trials is due to internal noise which is modeled by an additive noise having a standard model density function with mean 0 and standard deviation 1. The heterogeneous path is dominant whenever the net motion strength in the direction of the heterogeneous motion path (after adding noise) is positive (decision stage).

The energy-channel

In this section, we derive the characteristics of the single energy-channel. This energy-channel consists of two stages. The first stage is the nonlinear transformation (texture grabber). The simplest version of a texture grabber is a spatiotemporal linear filter followed by rectification (see Chubb & Sperling, 1989a, b). The output of this first stage (the texture activity) is fed into the second stage: motion energy analysis. Stages one and two are sketched in Fig. 10.

Stage 1: texture grabbers. It is now well-established (see review by Shapley & Enroth-Cugell, 1984), that early retinal gain-control mechanisms pass not stimulus luminance, but rather a signal approximating stimulus contrast, the normalized deviation of stimulus luminance from its local average. We assume that the spatiotemporal filters of Stage 1 operate on stimulus contrast.

The output magnitude of these filters varies over the visual field, depending on what textures happen to populate these regions. The output of a linear filter to a texture is variable and depends on the local phase of the texture. The purpose of rectification is to transform regions of highly variable response into regions of high average value, thus insuring that the rectified output registers the presence or absence of texture, independent of phase. Examples of rectification are half-wave rectification (setting negative values to zero) and full-wave rectification (anything that is symmetric with respect to input sign, such as absolute value or squaring).

The output of Stage 1 is called activity. The resulting transformation (accomplished by Stage 1) yields a spatiotemporal function whose value reflects the local texture preferences of the Stage 1 filter in the visual field as a function of time (see also Bergen & Adelson, 1988; Caelli, 1985). The activity transformation of the texture grabber depends on the amplitude *m* and spatial frequency *ω* of the textures involved.

In Expt 3, we have shown that texture activity is linear in texture amplitude. This is accommodated by a spatial filter that is linear in stimulus contrast. We can further characterize the spatial filter characteristics by the amplitude of its Fourier transform: $F(\omega)$. We assume that rectification is an absolute value operation. Thus, after rectification, the activity transformation T is proportional to m and to $F(\omega)$:

$$T(m, \omega) = mF(\omega). \qquad (2)$$

This texture activity T is fed into the second (motion energy analysis) stage.

Stage 2: motion energy analysis. The second stage (motion energy analysis) is a coincidence detector: it computes the product of the delayed activity at Location 1 with the current activity at Location 2 (van Santen & Sperling, 1984). For the displays we use in our experiments, the output of the second stage corresponds to motion strength.

To simplify the computation in the model, we assume that the first-stage spatiotemporal filter is space–time separable. Indeed, space–time separability seems to be the rule in apparent motion (Burt & Sperling, 1981; van de Grind, Koenderink & van Doorn, 1986).* Given space–time separability, we can ignore the temporal component of filtering because temporal patterns were not varied in our stimuli.

We proceed as follows. The perceived direction of motion is considered to be the outcome of a competition in motion strength between motion paths. Within a path the strength of motion between a patch of texture v and a patch of texture s is determined by the product of the activities of the first stage. We assume that the strengths of detectors for all paths are additive in the final motion percept, and adopt a linear combination model (Dosher, Sperling & Wurst, 1986). Additive internal noise determines the shape of the psychometric functions for motion direction as a function of amplitude.

Consider the strength model with respect to competition Scheme I (Fig. 3). In one direction there is a homogeneous motion path containing patches of identical texture s. In the opposite direction, there is a heterogeneous motion path containing patches of different textures s and v. For sine wave stimuli, a half-Reichardt model (simple product) is equivalent to the whole Reichardt model (difference of products) (van Santen & Sperling, 1985), so we need to consider just a simple product rule.

The strength of the heterogeneous motion path is:

$$S_{1,\text{he}}(m_v, \omega_v, m_s, \omega_s) = m_v F(\omega_v) m_s F(\omega_s). \qquad (3)$$

The motion strength $S_{1,\text{ho}}$ for the homogeneous motion path is equal to:

$$S_{1,\text{ho}}(m_s, \omega_s) = -m_s^2 F^2(\omega_s) \qquad (4)$$

(strength in the opposite direction has opposite sign).

Linear combination of both components with equal weights yields a net motion strength D_1 in the direction of the heterogeneous path:

$$D_1(m_v, \omega_v, m_s, \omega_s) = S_{1,\text{he}}(m_v, \omega_v, m_s, \omega_s) + S_{1,\text{ho}}(m_s, \omega_s). \qquad (5)$$

Response variability across trials is due to additive internal noise which is assumed to be distributed as a standard normal density function with mean 0 and standard deviation λ (Fig. 10). A linear addition of noise yields the internal decision variable i which has a normal distribution N with mean D and standard deviation λ.

According to signal detection theory (Green & Swets, 1966) the probability P of heterogeneous motion dominance is:

$$P_1(m_v; \omega_v) = P(i > 0) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N\{D_1(m_v, \omega_v, m_s, \omega_s), \lambda\}\, di. \qquad (6)$$

Substituting motion strengths [expressions (3) and (4)] into the additive linear combination [expression (5)] and then substituting [expression (5)] into the noise-driven decision process [expression (6)] yields:

$$P_1(m_v; \omega_v) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N\{[m_v F(\omega_v) m_s F(\omega_s) - m_s^2 F^2(\omega_s)], \lambda\}\, di. \qquad (7)$$

for the probability of heterogeneous motion dominance for Scheme I (Fig. 3).

Similar reasoning yields the net motion strength D_2 and the probability $P_2(m_v; \omega_v)$ of heterogeneous motion dominance in Scheme II (see Fig. 4):

$$D_2(m_v, \omega_v, m_s, \omega_s) = S_{2,\text{he}}(m_v, \omega_v, m_s, \omega_s) + S_{2,\text{ho}}(m_v, \omega_v) \qquad (8)$$

$$= m_v F(\omega_v) m_s F(\omega_s) - m_v^2 F^2(\omega_v) \qquad (9)$$

and

$$P_2(m_v; \omega_v) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N\{[m_v F(\omega_v) m_s F(\omega_s) - m_v^2 F^2(\omega_v)], \lambda\}\, di. \qquad (10)$$

This model predicts the transition and steepness at transitions of the probability curves for both the experiments.

Predictions for Scheme I

For different spatial frequencies ω_v of texture v, we measured the probability $P_1(m_v; \omega_v)$ of heterogeneous motion dominance as a function of the amplitude m_v of texture v. Our model predicts that the probability P_1 of heterogeneous motion dominance is an error function of the net motion strength D_1 [see equation (6)]. In this experiment, the net motion strength D_1 is linear in m_v. Hence, we expect an error function for the probability function $P_1(m_v; \omega_v)$ as a function of m_v [see equation (7)].

*It is reasonable to consider that the linear filter in the texture grabber may itself be composed as a weighted sum of many filters, i.e. filters that also are in the processing path for first-order motion. A linear filter composed as the sum of component filters would be space–time separable if each of its component filters were space–time separable and had the same temporal function, independent of spatial scale. This seems to be the case in motion processing (Burt & Sperling, 1981; van de Grind *et al.*, 1986).

Transition amplitude. The transition amplitude $\mu_1(\omega_v)$ is defined as the amplitude m_v of texture v at which the probability of heterogeneous motion dominance $P_1(m_v;\omega_v)$ is 50% for a given spatial frequency ω_v of texture v. Hence, for $m_v = \mu_1(\omega_v)$, the strength of the heterogeneous and homogeneous motion paths are balanced and we have $S_{1,he} = -S_{1,ho}$ or [see expressions (3) and (4)]:

$$\mu_1(\omega_v)F(\omega_v) = m_s F(\omega_s) = \kappa, \qquad (11)$$

where κ is a constant equal to the activity of standard texture *s*. If $F(\omega_v)$ is a low-pass filter, $\mu_1(\omega_v)$ will be a monotonically increasing function of ω_v (as supported by our experiments):

$$\mu_1(\omega_v) = \kappa F^{-1}(\omega_s). \qquad (12)$$

Steepness. The steepness $\sigma_1(\omega_v)$ is defined as the derivative of $P_1(m_v;\omega_v)$ with respect to m_v at transition amplitude $\mu_1(\omega_v)$:

$$\sigma_1(\omega_v) = \frac{\partial}{\partial m_v} P_1(m_v;\omega_v)|_{m_v = \mu_1(\omega_v)}$$
$$= \frac{1}{\sqrt{2\pi\lambda^2}}\kappa F(\omega_v). \qquad (13)$$

Thus, the steepness $\sigma_1(\omega_v)$ is expected to decrease as a function of the spatial frequency ω_v for low-pass filters (as supported by our experiments).

In conclusion we expect error functions for the probability $P_1(m_v;\omega_v)$ of heterogeneous motion dominance as a function of amplitude m_v with (a) a transition amplitude $\mu_1(\omega_v)$ that is inversely proportional with $F(\omega_v)$ and (b) a steepness $\sigma_1(\omega_v)$ that is proportional with $F(\omega_v)$. If we have low-pass filters, $F(\omega_v)$ decreases monotonically with spatial frequency ω_v.

Predictions for Scheme II

For different spatial frequencies ω_v of texture v, we measured the probability $P_2(m_v;\omega_v)$ of heterogeneous motion dominance as a function of the amplitude m_v of texture v. $P_2(m_v;\omega_v)$ is an error function of D_{II} [see equation (10)]. However, for Scheme II (unlike for Scheme I) D_{II} is not linear with the varied amplitude m_v of texture v. As we increase the amplitude m_v of texture v, D_{II} shows a quadratic dependence on m_v. Therefore, we do not expect an error function for $P_2(m_v;\omega_v)$.

If amplitude m_v of texture v is zero, the probability of heterogeneous motion dominance P_2 will be 50% (the motion stimulus is purely ambiguous!). Starting at $m_v = 0$, it first increases linearly with m_v, is maximal for $m_v = m_s F(\omega_s)/[2F(\omega_v)]$, and decreases again with further increases of m_v. Obviously, there may exist an amplitude $m_v = \mu_2$ (between the "optimal" amplitude, that yields a maximal D_2, and a very high amplitude, that yields a negative D_2) for which $P_2 = 50\%$.

Analogous to the derivation in the previous section, one can find the analytic expressions for the transition $\mu_2(\omega_v)$ and steepness $\sigma_2(\omega_v)$ of the probability curves for Scheme II. The expressions for the transition amplitudes are equal: $\mu_2(\omega_v) = \mu_1(\omega_v)$. The expressions for the steepness of the transitions for Scheme I and II differ only in sign: $\sigma_2(\omega_v) = -\sigma_1(\omega_v)$.

The texture grabber

We can simply find the Fourier transform $F(\omega)$ of the low-pass filter from the reciprocal transition $\mu_i^{-1}(\omega_v)$ [see expression (12)] and from the steepness $\sigma_i(\omega_v)$ as a function of spatial frequency ω_v [see expression (13)].

The reciprocal transition amplitudes are expected to be proportional to the function $F(\omega_v)$. Estimates of the reciprocal transition amplitudes $\mu_i^{-1}(\omega_v)$ are shown in Fig. 11.

From the reciprocal transitions in Fig. 11, it follows that $F(\omega)$ is a low-pass filter in the range of frequencies examined.

The model predicts that the steepness of the probability function is proportional with the function $F(\omega_v)$ and inversely proportional with λ (the strength of the internal noise). Thus, unlike the transition amplitude, the steepness is biased by the internal noise contribution. If the relative strength is constant and independent of the spatial frequency and amplitude of the patches of texture

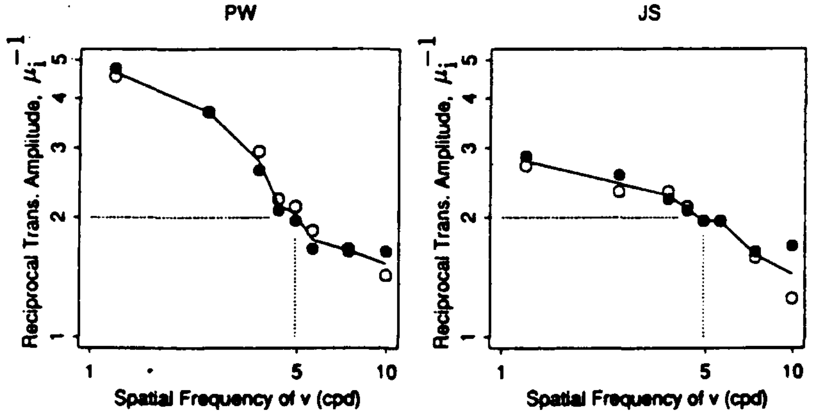


FIGURE 11. Reciprocal transitions $\mu_i^{-1}(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I; solid circles for Scheme II. The vertical dashed guide line indicates the spatial frequency of texture s: ω_s = 4.9 c/deg. The horizontal dashed guide line indicates the reciprocal amplitude of texture s. The solid line curve is the mean of the reciprocal transitions. In terms of the model, this curve shows the amplitude of the Fourier transform of the spatial filter $F(\omega)$ of the texture grabber involved [see equation (2)].

involved, the steepness $\sigma_i(\omega_v)$ is expected to be proportional with $F(\omega_v)$. Estimates of $\sigma_i(\omega_v)$ are shown in Fig. 8. The steepness shows a tendency to decrease with increasing spatial frequency. However, we find some nonmonotonicity, in particular for higher spatial frequencies. This may reflect a certain variability of the internal noise for different spatial frequencies.

EXPERIMENT 4: PERCEIVED CONTRAST

We have discussed texture grabbers and motion energy analysis in terms of objective amplitude of patches of texture. The experiments implied that the activity of the texture grabber increases monotonically with objective amplitude and decreases monotonically with spatial frequency. An interesting question is whether this relation is consistent with the subjective amplitude of static grating contrast as a function of spatial frequency. In other words, is the activity of a texture grabber simply proportional to the subjective amplitude?

To answer this question, we performed an amplitude discrimination experiment.

Method

In a two interval presentation subjects looked at an annulus containing either gratings s or v. In one interval we showed an annulus of gratings s (see frame f_2 of Fig. 3), with fixed amplitude $m_s = 0.5$ and fixed spatial frequency $\omega_s = 4.9$ c/deg. In the other interval we showed an annulus of gratings v (see frame f_2 of Fig. 4). with amplitude m_v and spatial frequency ω_v. The order of presentation of the intervals was randomized. Each annulus was shown for 133 msec (which is equal to the frame display time in the motion stimulus). The intervals were separated by a time interval of 133 msec in which the screen was uniform with background luminance. Apparatus, viewing conditions, and other aspects were identical to the motion experiment.

Procedure

The task of the subject was to indicate the interval that contained the patches of grating with the highest amplitude. We measured the probability $P_c(m_v; \omega_v)$ that observers judge the grating v as the grating with the highest amplitude as a function of the objective amplitude m_v of grating v. In the amplitude matching experiment, we examined two spatial frequencies: $\omega_v = 1.2$ c/deg, and $\omega_v = 7.4$ c/deg of grating v. These were the lowest and highest spatial frequencies for which we found transition invariance in our modern experiment. From these probability curves, we estimated the matching amplitude of grating v for which the perceived amplitude of grating s and v was equal. The precise estimation of the matching amplitude was analogous to the estimation of transition amplitude in the motion competition experiments.

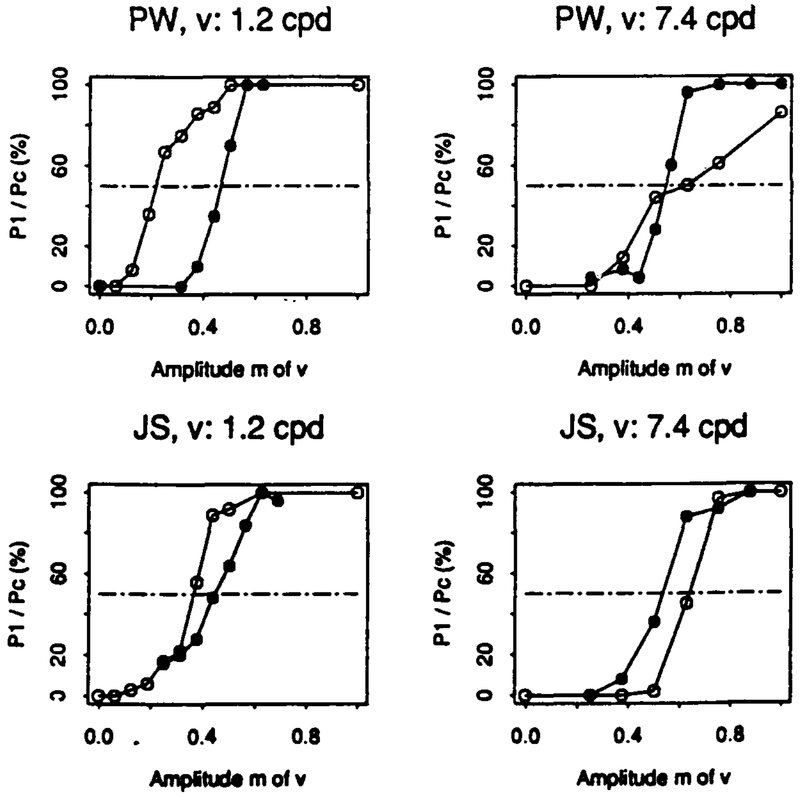


FIGURE 12. Results of the perceived amplitude experiment. Observers compared the amplitude of a grating v (spatial frequency ω_v and amplitude m_v) with the amplitude of texture s ($m_s = 0.5$, $\omega_s = 4.9$ c/deg). Shown are the probabilities P_c for judging the amplitude of v higher than that of s (solid circles). The matching amplitude for texture v is the crossing of the curve with the dashed 50% line. To compare the matching amplitude with the transition amplitude in the motion experiment, we have shown the probabilities $P_1(m_v)$ for Scheme I (open circles).

Results

In Fig. 12, we show the probabilities of judging the amplitude of grating v higher than that of grating s (with $m_s = 0.5$) as a function of objective amplitude m_v (solid circles). For all conditions and subjects, the perceived amplitude of texture v increases monotonically with its objective amplitude m_v. The amplitude m_v where the curve crosses the 50% guide line is the matching amplitude. For a "low" spatial frequency grating v ($\omega_v = 1.2$ c/deg), we find that the perceived amplitudes of s and v are matched when $m_v = 0.47$ for subject PW and $m_v = 0.44$ for JS. This matching amplitude is close to the objective amplitude $m_s = 0.5$ of grating s. For a "high" spatial frequency grating v ($\omega_v = 7.4$ c/deg), the matching amplitudes are $m_v = 0.54$ for PW and M_v 0.53 for JS.

The comparison of the matching amplitude with the transition amplitude in the motion experiments, we have also shown the probabilities to perceive heterogeneous motion using Scheme I as a function of m_v in the corresponding panels.

Discussion

Interestingly, the matching amplitudes for low and high spatial frequency gratings are approximately equal to the objective amplitude of grating s, for the range of amplitudes and spatial frequencies of grating v examined. That is, perceived amplitude does not depend on spatial frequency. However, the amplitude of grating v for balancing the motion paths when $\omega_v = 1.2$ c/deg for Scheme I was: $m_v = 0.22$ for subject PW and $m_v = 0.36$ for JS. Obviously, at the transition amplitude for the motion experiment, the perceived amplitude of grating s and v are markedly different. That is, the activities of the grating v are matched even when *both* spatial frequency and perceived amplitude are different from grating s. In conclusion, activity cannot be a function that depends solely on perceived amplitude.

EXPERIMENT 5: DICHOPTIC PRESENTATIONS

Motivation

We have successfully modeled the strength of motion-from-texture in terms of a texture grabber followed by motion energy analysis. Motion energy analysis is a type of motion computation that is *not* sensitive to correspondences in textural features. An interesting property of first-order motion energy analysis is that the neural substrate for such a process is organized so as to require successive stimulation to the same eye. When monocular motion information is not available to the observer first-order motion energy analysis fails.

The motion system that extracts first-order motion information of both eyes (when motion is presented dichoptically) has been classified as a correspondence-channel. For example, Pantle and Picciano (1976) studied apparent motion with a three-dot stimulus and reported element movement for monocular and binocular presentation, but group movement for dichoptic presentation. The group movement suggests a representation of features or shapes precedes the extraction of motion. Also, Georgeson and Shackleton (1989) show that drifting squarewave gratings with missing fundamental (MF) moved backwards while presented monocularly (following the third harmonic) but moved forwards when presented dichoptically. They suggested that the perceived direction of dichoptic apparent motion was consistent with a system that combines information across spatial frequency channels to identify local features and then tracks the location of corresponding features over time.

Generalizing the above reasoning to second-order motion, the motion mechanism for dichoptic presentations of our (second-order) stimuli would be sensitive to the similarity of the textures involved. Thus, the contribution of what we call correspondence-channels might be more pronounced when our competition schemes are presented dichoptically (sofar viewing has been binocular in our experiments). We tested our energy-channel model for motion-from-texture for both dichoptical and monocular presentations of our motion stimuli. This test may also locate the motion extraction process involved in our stimuli in terms of different levels in the visual nervous system (before or after the sites of binocular combination).

Results

The ambiguous motion competition Schemes I and II can be presented dichoptically in two different modes. In the first mode, the odd frames are presented in one eye and the even frames in the other. In this way, the spatiotemporal stimulus is purely ambiguous in each eye. Both the heterogeneous and the homogeneous paths are processed by dichoptic mechanisms. In this mode, dichoptic mechanisms are not competing with monocular mechanisms.

In the second mode, the patches of one texture type are presented in one eye and the patches of the second type of texture type are presented in one eye and the patches of the second type of texture in the other eye. In this way the homogeneous motion path (textures s for Scheme I) is presented in one eye, while the textures v in the other eye form a purely ambiguous stimulus. In this mode, dichoptic mechanisms processing the heterogeneous path have to compete with monocular mechanisms processing the homogeneous path.

We determined the psychometric functions for both competition schemes for a condition where the texture s and v differ two octaves in spatial frequency ($\omega_s = 4.9$ c/deg and $\omega_v = 1.2$ c/deg) for subject PW. The binocular results were presented in top-left panel of Fig. 6. As discussed for Expts 1 and 2, a difference between the transition amplitudes μ_1 and μ_2 indicates the involvement of additional (correspondence) channels. The results for monocular presentation were identical (within measurement error) to the results for binocular presentation. For both conditions, we find transition invariance: $\mu_1 = \mu_2 \approx 0.2$.

The results for both modes of dichoptic presentation were very similar to those for binocular presentation. That is, dichoptic presentation yields psychometric functions for Schemes I and II similar to those for binocular presentation. For adequate amplitude m_v heterogeneous motion dominated homogeneous motion for both modes of dichoptic presentation suggesting the dominance of an energy-channel even when monocular motion information was absent. However, the contribution of a correspondence-channel is noticeable for dichoptic presentations: transition invariance no longer holds. We found $\mu_1 \approx 0.2$ and $\mu_2 \approx 0.1$ for both modes of dichoptic presentation.

Discussion

Motion perception between patches of nonsimilar texture is easily perceived for both modes of dichoptic presentation (as predicted by our energy-channel). Even in the second mode, where a dichoptic heterogeneous motion path competes with a monocular homogeneous path, heterogeneous motion can easily dominate for small amplitude of texture v (e.g. $m_v > 0.2$ for Scheme I). These results suggest that dichoptic processing of our motion stimuli is dominated by the same mechanisms as monocular processing and that motion strength is not predicted by the similarity between textural features such as spatial frequency.

However, although dichoptic presentation leaves transition amplitude μ_1 for Scheme I unaffected, transition μ_2 for Scheme II decreases. This difference from the binocular results indicates a significant contribution of other channels when monocular information for the heterogeneous path is ambiguous. A more detailed investigation might be useful.

GENERAL DISCUSSION

Fallacy of correspondence matching

The experiments presented in this paper provide cogent evidence that texture similarity is not relevant to the texture-defined motion computation (within the range of spatiotemporal parameters varied in this experiment). As an example it was shown that motion between patches of texture that differ by two octaves in spatial frequency and a factor of 2 in amplitude can be stronger than motion between patches of identical texture.

The correspondence matching metaphor to explain visual processes in several visual domains seems to have lost predictive power. Correspondence matching fails to explain the dominance of (1) heterogeneous motion paths composed of textures that differ in spatial frequency and amplitude (this paper), (2) heterogeneous motion paths composed of elements that differ in size, orientation and luminance (Werkhoven *et al.*, 1990a, b), and (3) stereoscopic matches between elements that differ in size and luminance (Gulick & Lawson, 1976).

The visual motion system does not seem to be designed to establish correspondence between similar features in a motion sequence. This should not come as a surprise given the inherent difficulties in designing correspondence matching mechanisms. Such mechanisms would look for "similar features" in "successive" time samples of the spatiotemporal stimulus. However, what constitutes a feature, and how strict should similarity be taken?

Recently developed stimulus (motion) energy models for motion extraction bypass the correspondence problem and are more likely candidates for the kind of visual processing early in the visual system (Adelson & Bergen, 1985; Heeger, 1992). The energy-channel described in this paper is equivalent to such a motion energy computation, applied to a nonlinear transformation of the stimulus (van Santen & Sperling, 1984).

Contrast and motion

In Expt 3, we showed that the transition amplitude of texture v needed to balance the motion path s, v with the motion path s, s varies linearly with the amplitude of texture s. In the context of our model, this means that the activity of a texture grabber is approximately linear in texture amplitude. In fact, we find linearity even for high amplitudes in the range of 50–100%. As a consequence of this amplitude linearity, motion strength varies linearly with the amplitude of each of the texture inputs. That is, the strength of motion between two textures with identical texture amplitude is quadratic with this amplitude. Approximate amplitude linearity of the input lines for first-order motion energy analysis was also found for experiments with spatiotemporal modulations of luminance Werkhoven *et al.* (1990b).

It should be noted, that the linear amplitude dependency is at odds with the amplitude thresholds for motion direction discrimination reported by Nakayama and Silverman (1985). They measured the smallest phase shift (yielding threshold direction discrimination performance) of sinusoidal gratings as a function of grating amplitude. The smallest phase shift yielding threshold performance leveled off for grating amplitudes exceeding 5%. They interpreted their finding in terms of a amplitude saturation function. However, their results are open to a different interpretation in which the minimum phase shift is limited by other (spatial) properties of the motion extraction mechanism leaving the amplitude dependency unknown.

A shared motion analysis stage?

An intriguing question is how mechanisms for the extraction of motion carried by the spatiotemporal modulation of luminance relate to those for extracting motion carried by the spatiotemporal modulation of texture type. To discriminate both mechanisms we have to compare the characteristics of the perception of both motion types. For example, Turano and Pantle (1989) studied velocity discrimination performance for both types of motion stimuli and showed similar discrimination characteristics. Their results support the hypothesis of a higher order (motion analysis) mechanism that accepts input from both the luminance domain as well as texture domain.

A shared motion energy analysis stage for the two types of motion is also supported by our finding that strength of motion-from-texture is ruled by the same metric as motion in the luminance domain. Motion strength is the covariance (or product) of local activities. This activity is simply the luminance itself when the motion is carried by luminance (van Santen & Sperling, 1984) or a nonlinear transformation of the luminance pattern for motion-from-texture (this paper).

In conclusion, the extraction of motion from the spatiotemporal modulations of luminance and that of texture types seems to be mediated by a shared motion energy analysis stage. However, additional experiments with different paradigms may weaken this idea. For example, Mather (1991) showed that both motion types produce motion after effects, but that the duration of the aftereffects were significantly different.

Transitivity and additivity

Under the assumption of energy channels and channel summation, the transition invariance of a pair of textures s and v implies that s and v are (texture-defined) motion metamers. That is, all such textures v in this metameric class yield identical motion strength when embedded in a motion path s, v.

Metamery yields two strong predictions. First, metamery predicts transitivity: if textures a and b are metameric with s, then a is metameric with b. Second, metamery predicts additivity: if textures a and b are metameric with s, then any linear combination $\alpha a + \beta b$ (with $\alpha + \beta = 1$) is metameric with s.

These predictions have not yet been tested.

Motion transparency

The energy-channel proposed in this paper computes the difference between left- and rightward motion. This implies that motion transparency (the simultaneous detection of left- and rightward motion) is not readily accommodated in this model. Because the motion analysis component of the energy-channel is a Reichardt-correlator, the motion energy of the left- and rightward motion path are no explicit intermediate results). However, occasionally, observers reported transparency for stimuli that were nearly balanced.

Adelson and Bergen (1985) addressed this issue by pointing out that although their energy detector was functionally equivalent to correlation detector, the intermediate results are not. Specifically, the energy of left and rightward motion are explicit intermediate results in energy detectors, but not in correlation detectors (the output of a half Reichardt-correlation is the half-phase opponent energy!). Although our conclusions do not depend on the specific choice of motion model, a further study of transparency in this context might reveal the specific type of detector involved.

Extension of the parameter space

It is important to remember that we have shown the one-dimensionality of the motion-from-texture computation only with respect to parallel sinewave patches that differ in spatial frequency and amplitude. Chubb and Sperling (1991) found that motion-from-texture could be carried by differences in spatial orientation, although differences in orientation did not produce as vigorous motion as did differences in spatial frequency. This observation indicates that orientation (and possibly other properties) are relevant to motion-from-texture. It would be interesting to determine the dimensionality of the computation for a larger class of stimuli.

Although motion strength at a "frame time" τ of 8/60 sec is exclusively determined by the product of activities, we can not exclude that effects of texture similarity are stronger at longer frame time. In fact, the temporal frequency of texture modulation in our experiments is 1.9 Hz (one cycle consists of four frames of 133 msec each). At slower temporal frequencies, the processing time for the textures increases, perhaps enabling more elaborate "texture grabber" filters or correspondence-channels to contribute to motion strength.

Effects of other properties (e.g. orientation) and temporal parameters are currently under investigation.

REFERENCES

Adelson, E. H. & Bergen, J. R. (1985). Spatio-temporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2,* 284–299.

Adelson, E. H. & Bergen, J. R. (1986). The extraction of spatio-temporal energy in human and machine vision. In *Proceedings: Workshop on motion: Representation and analysis* (pp. 151–155) IEEE Computer Society Press.

Balliet, R. & Nakayama, K. (1978). Training of voluntary torsion. *Investigative Ophthalmology and Visual Science, 17,* 303–314.

Bergen, J. R. & Adelson, E. H. (1988). Early vision and texture perception. *Nature, 333,* 363–364.

Braddick, O. J. (1980). Low-level and high-level processes in apparent movement. *Philosophical Transactions of the Royal Society of London B, 290,* 137–151.

Burt, P. & Sperling, G. (1981). Time, distance and feature trade-offs in visual apparent motion. *Psychological Review, 88,* 171–195.

Caelli, T. (1985). Three processing characteristics of visual texture segregation. *Spatial Vision, 1,* 19–30.

Cavanagh, P. & Mather, G. (1989). Motion: The long and the short of it. *Spatial Vision, 4,* 103–129.

Cavanagh, P., Arguin, M. & von Grünau, M. (1989). Interattribute apparent motion. *Vision Research, 29,* 1197–1204.

Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, 5,* 1986–2007.

Chubb, C. & Sperling, G. (1989a). Second-order motion perception: Space/time separable mechanisms. *Proceedings: Workshop on visual motion, Irvine, California, 1989* (pp. 126–138). IEEE Computer Society Press.

Chubb, C. & Sperling, G. (1989b). Two motion perception mechanisms revealed by distance driven reversal of apparent motion. Proceedings: Workshop on visual motion, Irvine, California, 1989. *Proceedings of the National Academy of Sciences U.S.A., 86,* 2985–2989.

Chubb, C. & Sperling, G. (1991). Texture quilts: Basic tools for studying motion-from-texture. *Journal of Mathematical Psychology, 35,* 411–442.

Dosher, B. A., Sperling, G. & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26,* 973–990.

Farrell, J. E., Pavel, M. & Sperling, G. (1990). The visible persistence of stimuli in stroboscopic motion. *Vision Research, 30,* 921–936.

Georgeson, M. A. & Shackleton, T. M. (1989). Monocular motion sensing, binocular motion perception. *Vision Research, 29,* 1511–1523.

Graham, N. (1992). Complex channels, early local nonlinearities, and normalization in texture segregation. In Landy, M. S. & Movshon,

J. A. (Eds), *Computational models of visual processing* (Chap. 18). Cambridge, Mass.: MIT Press.

Green, M. (1986). What determines correspondence strength in apparent motion? *Vision Research, 26,* 599–607.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

van de Grind, W. A. Koenderink, J. J. & van Doorn, A. J. (1986). The distribution of human motion detector properties in the monocular visual field. *Vision Research, 26,* 797–810.

Gulick, W. L. & Lawson, R. B. (1976). *Human stereopsis: A psychophysical analysis.* New York: Oxford University Press.

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A, 4,* 1455–1471.

Heeger, D. J. (1992). Nonlinear model of neural responses in cat visual cortex. In Landy, M. S. & Movshon, J. A. (Eds), *Computational models of visual processing* (Chap. 9). Cambridge, Mass.: MIT Press.

Kolers, P. A. (1972). *Aspects of motion perception.* Oxford: Pergamon Press.

Lelkens, A. M. M. & Koenderink, J. J. (1984). Illusory motion in visual displays. *Vision Research, 24,* 1083–1090.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London, 200,* 269–294.

Mather, G. (1991). First-order and second-order visual processes in the perception of motion and tilt. *Vision Research, 31,* 161–167.

Moulden, B. & Begg, H. (1986). Some tests of the Marr-Ullman model of movement detection. *Perception, 15,* 139–155.

Nakayama, K. & Silverman, G. H. (1985). Detection and discrimination of sinusoidal grating displacements. *Journal of the Optical Society of America A, 2,* 267–274.

Navon, D. (1976). Irrelevance of figural identify for resolving ambiguities in apparent motion. *Journal of Experimental Psychology, Human Perception and Performance, 2,* 130–138.

Pantle, A. & Picciano, L. (1976). A multistable movement display: Evidence for two separate motion systems in human vision. *Science, 193,* 500–502.

Papathomas, T. V., Gorea, A. & Julesz, B. (1991). Two carriers for motion perception: Color and luminance. *Vision Research, 31,* 1883–1891.

Ramachandran, V. S., Rao, M. V. & Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research, 13,* 1399–1401.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Rosenblith, W. A. (Ed.), *Sensory communication.* New York: Wiley.

Robson, J. G. (1980). *Neural images: The physiological basis of spatial vision.* In Harris, C. (Ed.), *Visual coding and adaptability* (pp. 177–214). Hillsdale, N. J.: Erlbaum.

van Santen, J. P. H. & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, 1,* 451–473.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2,* 300–321.

Shapley, R. & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research, B3,* 263–346.

Shechter, S., Hochstein, S. & Hillman, P. (1989a). Size, flux and luminance effects in the apparent motion correspondence process. *Vision Research, 29,* 579–591.

Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods and Instrumentation, 8,* 144–151.

Turano, K. & Pantle, A. (1989). On the mechanism that encodes the movement of contrast variations: velocity discrimination. *Vision Research, 29,* 207–221.

Ullman, S. (1980). The effect of similarity between bar segments on the correspondence strength in apparent motion. *Perception, 9,* 617–626.

Victor, J. D. & Conte, M. M. (1990). Motion mechanisms have only limited access to form information. *Vision Research, 30,* 289–301.

Watson, A. B. (1986). Apparent motion occurs only between similar spatial frequencies. *Vision Research, 26,* 1727–1730.

Werkhoven, P. & Koenderink, J. J. (1991). Reversed rotary motion perception. *Journal of the Optical Society of America A, 8,* 1510–1516.

Werkhoven, P., Snippe, H. P. & Koenderink, J. J. (1990a). Effects of element orientation on apparent motion perception. *Perception and Psychophysics, 47,* 509–525.

Werkhoven, P., Snippe, H. P. & Koenderink, J. J. (1990b). Metrics for the strength of low level motion perception. *Journal of Visual Communication and Image Representation, 1,* 176–188.

Acknowledgement—This work was supported by the AFOSR Grant 91-0178.

APPENDIX

Multiple Energy-Channels and Transition Invariance

A system of multiple energy-channels

We propose a multi-channel model (multiple energy-channels) for computing the strength of motion-from-texture. The model consists of two stages, as shown in Fig. 13.

Stimulus transformation: texture grabbers. Stage 1 consists of n types of texture grabbers—where each type of texture grabber i is described by nonlinear spatiotemporal transformations T_i, $i = 1 \ldots n$, of the optical input. Each transformation yields a spatiotemporal function $T_i(\varphi, t)$ whose value reflects the local texture preferences of the Stage 1 filters in the visual field as a function of position φ and time t. (We use φ for position because, in our essentially one-dimensional stimulus, the texture position is determined by the angle φ.) The output of these texture grabbers is called activity. The n different transformations T_i of Stage 1 transform the optical input into n activity representations.

Motion detection. Stage 2 is a set of motion detectors. For specificity, but without loss of generality (see van Santen & Sperling, 1984; Chubb & Sperling, 1988, 1991) we adopt Reichardt's scheme for standard motion analysis (Reichardt, 1961) which consists of two oppositely tuned coincidence detectors. Motion detectors operate on the outputs of the texture grabbers. Each type of texture grabber (transformation T_i) has its own, unique set of motion detectors. A transformation T_i together with its motion detectors is called a motion channel i.

A coincidence detector performs a multiplication operation on the current activity $T_i(\varphi, t)$ at position φ at time t and the (delayed) activity $T_i(\varphi - \Delta\varphi, t - \Delta t)$ at position $\varphi - \Delta\varphi$ and time $t - \Delta t$. Hence, the output of the coincidence detector is: $T_i(\varphi - \Delta\varphi, t - \Delta t)T_i(\varphi, t)$. The outputs of two coincidence detectors

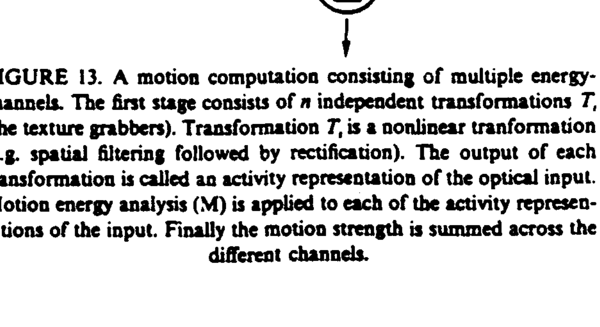


FIGURE 13. A motion computation consisting of multiple energy-channels. The first stage consists of n independent transformations T_i (the texture grabbers). Transformation T_i is a nonlinear tranformation (e.g. spatial filtering followed by rectification). The output of each transformation is called an activity representation of the optical input. Motion energy analysis (M) is applied to each of the activity representations of the input. Finally the motion strength is summed across the different channels.

tuned to identical velocities but opposite directions are subtracted to yield a net motion strength $D_i(\varphi, t)$:

$$D_i(\varphi, t) = T_i(\varphi - \Delta\varphi, t - \Delta t)T_i(\varphi, t) - T_i(\varphi - \Delta\varphi, t)T_i(\varphi, t - \Delta t). \tag{14}$$

Channel i has a positive output for motion in the direction of positive φ and a negative output for motion in the opposite direction.

Summation. In a one-dimensional motion computation, the outputs of a system of energy-channels described above (represented in an n dimensional channel space) are essentially mapped to a single (decision) dimension: the final net motion strength. This mapping maps an $(n - 1)$-dimensional manifold in the channel space to a single point in the one-dimensional decision space (final motion strength). For example, channel summation maps a planar surface in the channel space to zero final motion strength (for Scheme I). For other combination rules than summation, other (nonplanar) surfaces will map to zero final motion strength. However, when we assume that this mapping is continuous and differentiable, these true manifolds are in first order approximated by a planar surface for small channel signals at transition points. Channel summation is a sufficient first-order combination rule.

Summation of channels D_i yields net motion strength D:

$$D(\varphi, t) = \sum_{i=1}^{n} D_i(\varphi, t). \tag{15}$$

Predictions for competition schemes

We apply the multi-channel computation to competition Schemes I and II (see Figs 3 and 4). Consider first Scheme I. The heterogeneous path is the motion between texture s (at time $t - \Delta t$ and position $\varphi - \Delta\varphi$) and texture v (at time t and position φ). Let $T_{i,s}$ be the activity of texture grabber T_i for texture s, and $T_{i,v}$ the activity of texture grabber T_i for texture v. The output of channel i for this path is the product of the delayed activity $T_{i,s}$ of texture s and the current activity $T_{i,v}$ of texture v. For simplicity, we will use the vector notation:

$$\vec{T}_s = \begin{pmatrix} T_{1,s} \\ T_{2,s} \\ \cdot \\ \cdot \\ T_{n,s} \end{pmatrix} \text{ and } \vec{T}_v = \begin{pmatrix} T_{1,v} \\ T_{2,v} \\ \cdot \\ \cdot \\ T_{n,v} \end{pmatrix}. \tag{16}$$

The vectors $\vec{T}_s$ and $\vec{T}_v$ are the activity vectors of textures s and v respectively. An activity vector represents the activity of a texture in the n-dimensional transformation space (T-space) defined by transformations $T_1 \cdots T_n$.

For Scheme I, the motion strengths $S_{1,he}$ summed over all channels for the heterogeneous path can be written as the vector product:

$$S_{1,he} = \vec{T}_s \cdot \vec{T}_v = \sum_{i=1}^{n} T_{i,s}T_{i,v}. \tag{17}$$

We have arbitrarily assigned a positive sign to motion strength in this direction. Motion in the opposite direction has a negative sign [see equation (14)]. The output of channel i for the homogeneous path (between textures s) is the squared output of transformation $T_{i,s}$. The motion strength $S_{1,ho}$ of the homogeneous path is (after summing all channels) is:

$$S_{1,ho} = -\vec{T}_s \cdot \vec{T}_s. \tag{18}$$

Adding equations (6) and (7) gives the net motion strength D_1 in the direction of the heterogeneous path for Scheme I:

$$D_1 = \vec{T}_s \cdot (\vec{T}_v - \vec{T}_s). \tag{19}$$

Analogously, the net motion strength D_2 in the direction of the heterogeneous path for Scheme II is:

$$D_2 = \vec{T}_v \cdot (\vec{T}_s - \vec{T}_v). \tag{20}$$

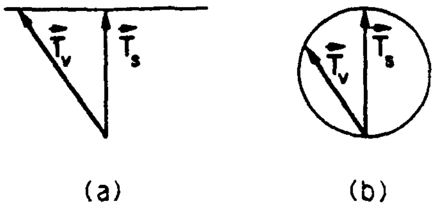


FIGURE 14. Solutions for transitions (path equality) in a two-dimensional T-space. Each texture in a motion path is processed by different texture grabbers. Vector $\vec{T}_v$ represents the activity of texture v in T-space, vector T_s that of s. The collection of activity vectors $\vec{T}_v$ that satisfy the constraints for path equality are given by the thin line in (a) for Scheme I and by a thin circle in (b) for Scheme II.

Transitions: Scheme I

At a transition for Scheme I, the net motion strength D_1 is zero

$$D_1 = \vec{T}_s \cdot (\vec{T}_v - \vec{T}_s) = 0. \tag{21}$$

There exists an $(n - 1)$-dimensional plane of $\vec{T}_v$ vectors in T-space for which the motion strength of the heterogeneous and homogeneous motion paths are balanced (the vectors $\vec{T}_v$ for which the difference vector $\vec{T}_v - \vec{T}_s$ are orthogonal to vector $\vec{T}_s$).

Consider, for example, a two-dimensional T-space (a two-channel motion computation). The vectors $\vec{T}_v$ in T-space that satisfy equation (21) for a certain vector $\vec{T}_s$ must end on the thin guide line in Fig. 14(a).

It should be noted in passing, that the net heterogeneous motion strength $D_1 = \vec{T}_s \cdot (\vec{T}_v - \vec{T}_s)$ can be positive. Hence, even in a multi-channel computation, the strength of the heterogeneous motion path can dominate.

Transitions: Scheme II

Similarly, at a transition for Scheme II (Fig. 4), the net motion strength D_2 is zero:

$$D_2 = \vec{T}_v \cdot (\vec{T}_s - \vec{T}_v) = 0. \tag{22}$$

The $(n - 1)$-dimensional solution of $\vec{T}_v$ vectors in T-space for which the motion strength of the heterogeneous and homogeneous motion paths are balanced is not a plane. For example, we consider again the two-dimensional T-space. The vectors $\vec{T}_v$ in T-space that satisfy equation (22) for a certain vector $\vec{T}_s$ end on a circle containing $\vec{T}_s$ [see Fig. 14(b)].

Transition invariance

Using only the result for Scheme I, we cannot discriminate between a single-channel $(n = 1)$ and multi-channel computations $(n > 1)$. either single- or multi-channel computations might yield solutions to equation (21). To resolve the issue, we need the constraint of transition invariance.

Transition invariance means that once the motion strength of the heterogeneous path and that of the homogeneous motion path are balanced for a particular pair of textures s and v for Scheme I, this balance is not disturbed by interchanging the textures s and v (yielding Scheme II). We now show that transition invariance is inconsistent with a multi-channel computation.

The transitions are invariant if the activity vector $\vec{T}_v$ simultaneously satisfies equations (21) and (22). Because the difference vector $\vec{T}_s - \vec{T}_v$ is always in the plane defined by vector $\vec{T}_s$ and vector $\vec{T}_v$, the only vector $\vec{T}_v$ that satisfies both equations is $\vec{T}_v = \vec{T}_s$.

Vector $\vec{T}_v$ is equal to vector $\vec{T}_s$ if each transformation T_i involved in the motion computation has an equal output for both textures v and s:

$$T_{i,s} = T_{i,v} \quad (i = 1 \cdots n). \tag{23}$$

Equation (23) represents a very strong constraint for the ensemble of transformations that might be involved in a multi-channel computation. Every transformation T_i must have an isoactivity contour as a function of all textural properties (e.g. frequency–amplitude space) that contains both the activity of texture s and that of texture v. Furthermore, transition invariance holds for different texture pairs (s, v); the iso-activity contours of each transformation T_i must be identical for all these pairs. Transformations that are identical at arbitrarily many observable points, are identical in the range of observable points. To say that all T_i are identical is equivalent to saying that there is only one T_i, that is, the T-space is one-dimensional.

George Sperling, Stephen A. Wurst and Zhong-Lin Lu. Using Repetion Detection to Define and Localize the Processes of Selective Attention. In Meyer and Kornblum, (Ed.) *Attention and Performance XIV*. Cambridge, Massachusetts: The MIT Press, 1992. Pp. 265-298.

12 Using Repetition Detection to Define and Localize the Processes of Selective Attention

George Sperling, Stephen A. Wurst, and Zhong-Lin Lu

12.1 INTRODUCTION

Overview

In our repetition-detection task, subjects search a rapid sequence of thirty frames for a stimulus that is repeated within four frames. Successful detection implies that a match occurs between an incoming item and a recent item retained in short-term visual repetition memory (STVRM).

We test selective attention to physical features in a single location within which successive items alternate in color, size, or spatial frequency. For example, in the size condition, large and small items strictly alternate, and subjects attend selectively *to small* (or to *large*) items. Selective attention to *small* facilitates detecting small-small repetitions and impairs detection of large-large repetitions (the benefit and cost of selective attention). In a control condition, the *large* items are replaced by blanks. The size of the attention benefit for small relative to the control performance gives the efficiency of attentional filtering relative to perfect optical filtering.

Whereas selective attention (relative to equal attention) facilitates homogeneous (e.g., small-small) repetition detections, it usually impairs heterogeneous detections (large-small or small-large). Comparisons of attention costs and benefits for homogeneous and for heterogeneous detections admit the following inferences: physical features are represented in STVRM; attentional filtering occurs before stimuli are recorded in STVRM; in some conditions, some subjects use strategies that encode the attention state of an item in STVRM.

Background: Early versus Late Selective Filtering

Theories of selective attention postulate that the human information processing system is limited in its capacity and that attention serves to select information to be processed from other, competing information (e.g., Broadbent 1958; Deutsch and Deutsch 1963; Norman 1968). Indeed, selective filtering of unattended information has been proposed as a mechanism in numerous visual processing tasks.

There is abundant evidence that selective attention can function as a mechanism to differentially filter information from different spatial locations (see reviews by Sperling and Dosher 1986; Sperling and Weichselgartner 1993). However, we find no convincing evidence that attention can function as a mechanism for selecting information on the basis of physical features when items containing different constellations of features occur at the same location. Rather, the data are consistent with a theory that asserts that stimulus features serve only to guide spatial attention. That is, whenever selection appears on the basis of the physical features of visual stimuli (such as color, spatial frequency content, size, etc.), these features serve to bring attention to a particular location, but the attentional filtering is on the basis of location rather than on the basis of feature. To test this theory, it is critical to present more information than can be successfully processed at a single location, and to observe whether, at this single location, attentional filtering is possible on the basis of physical features.

Selection from Streams

It is trivial to demonstrate that attentional filtering can occur within a given spatial location. Consider, for example, the following gedanken experiment. Subjects view a stream of alternating black and white digits on a gray background. Subjects are asked to compute the sum of the white digits and to ignore the black digits. Obviously, subjects can perform this task when the stream is slow enough, but this would not be profoundly revealing about selective attentional processes because we already know that selection can occur at a cognitive or a decision level of processing. The interesting questions about selective attention concern whether it can operate at an earlier sensory or perceptual level (reviewed in Sperling and Dosher 1986).

Search Procedures A useful technique for studying attentional selection at a single location is to present a rapid stream of items at a location too rapidly to permit all items to be processed perfectly. Attentional selection can then be used to determine which items are processed. There are a number of tasks that involve items that are presented in a rapid visual stream at a single location. For example, Sperling et al. (1971) studied rapid visual search as a function of the number of locations in which streams of items were presented. However, the problem with search experiments is that, so far, no procedure has been developed to determine whether attentional selection (i.e., rejection of nontarget items) occurs at the perceptual or at the decision level of processing. Indeed, recent theories of selective filtering (Cave and Wolfe 1990; Duncan and Humphreys 1989; Pavel 1991; Wright and Main 1991; cf. Hoffman 1979) propose various cue-weighting algorithms to determine the sequence of attentional selections in visual search. Such weighting processes are typical of decision processes, although the algorithms themselves are neutral with regard to whether they operate at a perceptual or a decision level of processing.

Feature-based Partial Reports from Streams Another task involving a stream is the selective recall of items according to their physical characteristics. The procedure involves the selection of items from a rapid stream according to whether or not the target items have a distinguishing characteristic such as a ring around them, or whether they are brighter than their neighbors. Subjects can extract single target items from a rapid stream (Intraub 1985; Weichselgartner and Sperling 1987), or even a short sequence of four targets (Weichselgartner 1984). In fact, such experiments are partial report experiments in which the many items (from among which a few are selected for a partial report) are arrayed in time rather than in space as in the more usual procedure (Sperling 1960).

Feature-based Partial Reports from Spatial Arrays In spatial arrays, subjects can select items for partial report that have a ring around them (Averbach and Sperling 1961) or items that merely are pointed at by a short bar marker—a minimal feature for selection. When subjects are required to report only items of a particular color from briefly exposed letter matrices, these partial reports are not much better than whole reports (von Wright 1968). Similarly, when subjects are required to report only digits from mixed arrays of letters and digits, subjects do not report more digits than when they must report both letters and digits (e.g., Sperling 1960). Both of these studies required subjects to extract both item-identity and location information from briefly exposed arrays. When subjects are required only to report the item identities and not locations, partial reports according to feature easily surpass whole reports (e.g., selecting solid from outline characters; Merikle 1980). Thus, with comparable response requirements, feature-cued partial reports are comparably successful in temporal streams and in spatial arrays.

Partial Reports according to Spatial or Purely Temporal (versus Featural) Cues Originally, partial reports were studied in spatial arrays, and the selection cue designated one of several rows of characters—purely spatial selection (e.g., Sperling 1960, 1963). With spatial cues, there is a large and consistent partial-report advantage. When subjects must use a temporal cue to make a partial-report selection of four items from a rapid temporal stream, item selection appears to be based on a temporal window of attention (Sperling and Reeves 1980; Reeves and Sperling 1986; Weichselgartner and Sperling 1987). The subject's temporal window for selection from temporal streams is perfectly analogous to the spatial window for selection from spatial arrays (e.g., LaBerge and Brown 1989).

The Locus of Feature-based Attentional Selection Partial-report paradigms primarily focus on the process whereby information is selected for inclusion in short-term memory. That feature-based attentional selection of information for partial reports can occur in streams and in arrays merely places the level of attentional selection below the level of short-term memory. This constraint is unremarkable. Therefore, it is search tasks that seem most often to have been called forth to resolve the issue of early versus late selection on the basis of physical features (recent examples include Nakayama and Silverman 1986; Neisser 1967; Treisman 1977; Treisman 1986; Treisman and Gelade 1980; see Folk and Egeth 1989 for a review). Closely related issues are automatic versus controlled processing (Shiffrin and Schneider 1977), speeded classification (e.g., Felfoldy and Garner 1971; Garner 1978) and auditory selective attention (Swets 1984). The ambiguity of current search theories concerning the level of attentional selection was noted above. This is not the place for a review and critique of the many other approaches to these problems in the visual and auditory domains. Instead, we offer new variations of a repetition-detection task and new analyses that are particularly well suited to defining the locus of feature-based attentional selection (i.e.. perceptual filtering according to physical properties).

Repetition-Detection Paradigm

The repetition-detection paradigm (Kaufman 1978; Wurst 1989; Sperling and Kaufman 1991) seems particularly well suited for the study of attentional selection based on physical features. In this paradigm (fig. 12.1), a stream of thirty digits is presented rapidly (typically, 9.1 digits per sec). Within this stream, every digit is repeated three times, but only one digit is repeated within four sequence positions (lag 4 or less); all other digits are repeated with lags of nine or more. The subject is instructed to detect the recently repeated digit. Successful performance of this task obviously depends on the subject's ability to match incoming digits with previously presented digits in memory. Because all digits are repeated exactly three times within a list, only memory that discriminates short-lag repetitions from long-lag repetitions is useful for performing this task.

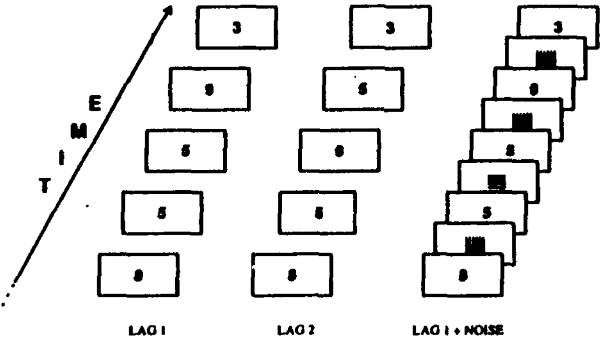


Figure 12.1 The repetition-detection paradigm. The leftmost sequence (lag 1) represents five consecutive frames from the middle of a longer sequence of frames. The target repetition is the digit 5. The middle sequence illustrates repetition of the digit 5 with lag 2. The rightmost sequence illustrates Kaufman's (1978) noise condition with lag 1. A grid of randomly chosen vertical or horizontal lines is interposed between each digit frame; repetition-detection performance was unimpaired.

In previous research (Kaufman 1978; Sperling and Kaufman 1991), it was found that, at lag 1, repetition detection was typically better than 80 percent correct, and that by lag 4 it had dropped below 30 or 40 percent. Adding a noise field between successive frames (fig. 12.1) did not impair performance, even when the noise field was so intense that, if it were simultaneous with digit presentations, it would have rendered them illegible. This immunity to visual masking suggests a central memory locus for short-term visual repetition memory (STVRM), even at lag 1.

In another adaptation of the task (Kaufman 1978; Sperling and Kaufman 1991), it was found that using nonsense shapes as stimuli instead of digits yielded equivalent results. This suggests that STVRM is visual rather than verbal or semantic.

Wurst (1989) used dicoptic presentations to demonstrate that the locus of short-term visual repetition memory (STVRM) was after the locus of binocular combination. A particularly interesting finding in Wurst's dicoptic viewing procedure was that one eye was given priority over the other eye. Thus, a *filtering* of items by the eye of presentation may have been occurring even though items were presented alternately (never simultaneously) to the two eyes and though, in control conditions, monocular performance was the same for both eyes. The present study was undertaken to determine whether selection could occur by varying stimulus attributes other than the eye of presentation.

Plan of the Experiments

To investigate the role of attention in the short-term visual repetition memory task, as in the previous studies, digits are presented in the same spatial location while being viewed binocularly. Two levels of a dimension are employed (e.g., large and small sizes of digits), and digits alternate between the levels. We will call a level of a dimension a *feature*. For example, *small* and *large* are features within the size dimension. In this study, five stimulus dimensions that have typically been employed in attention research (e.g., Nakayama and Silverman 1986; Treisman 1982; Sagi 1988)—size, angular orientation, spatial bandpass filtering, contrast polarity (black-on-gray vs. white-on-gray), and color—are examined separately. Additionally, we examine one feature pair (small black vs. large white). Digits with a different feature (e.g., large and small size) are alternated at the same location. We determine the ability of subjects to attend *selectively* to items with one feature (or feature pair) while ignoring items with the other feature (or feature pair).

12.2 METHOD

Experiment 1 examines five individual stimulus dimensions—orientation, size, contrast polarity, color, and spatial bandpass filtering—and one dimensional pair (small black vs. large white). Experiment 2 investigates three sets of stimulus characters and will be described more fully below.

A stimulus sequence consists of thirty consecutive digits. A position in the sequence is called a *frame*; thus we say the ith digit occurs in the *i*th frame. Stimuli in a sequence alternately exhibit one level *A* of a dimension on odd-numbered frames, and the other level *B* of the same dimension on even-numbered frames. We call such as sequence $\frac{1}{2}A + \frac{1}{2}B$. If subjects were completely successful in selectively filtering out unattended *B* stimuli on the even-numbered frames, detection of the repetitions of the attended-to-feature in a $\frac{1}{2}A + \frac{1}{2}B$ sequence would be similar to a control condition ($\frac{1}{2}A$) in which the even-numbered frames were simply blank. If the selection were totally unsuccessful, for example, if the features were indiscriminable, then the alternating feature sequence should be as difficult as a same-feature sequence (*A*). Consider performance in the two control conditions $\frac{1}{2}A$ and *A*. The point between these two performances where performance with $\frac{1}{2}A + \frac{1}{2}B$ falls indicates the success of attentional filtering. This is the broad plan of the experiments. Additional complications will become apparent as the story unfolds.

Stimulus Generation

Frames The repetition-detection procedure (Kaufman 1978; Sperling and Kaufman 1991), was used in this experiment. Each trial consisted of a stream of thirty digits displayed on a video monitor. A digit was painted three times (three refreshes), followed by six refreshes of a blank, gray screen, all at sixty refreshes per second. The sequence of nine refreshes (digit plus subsequent blank screen) is called a *frame*. The frame duration is 150 msec; equivalently, the digit-to-digit stimulus onset asynchrony (SOA) is 150 msec. A digit sequence was composed of thirty frames: the ten digits, each presented three times.

Lag To distinguish the different types of repetitions that occur, we use the term *lag*. When a digit occurs in frame *i* of the sequence, and then again in frame *j*, $1 \leq i < j \leq 30$, the digit is defined as being repeated with lag $j - i$ (see fig. 12.1). Only the target digit was repeated within a lag of 4 or less; all other repetitions of the digits were separated by eight or more intervening digits (lag ≥ 9). To generate a stimulus sequence, the first digit is chosen randomly. Subsequently, at any point in sequence generation, the requirement that no digit be repeated with lag ≤ 8 restricts the number of digits eligible to be chosen. At each point (except the critical repetition), the new digit was chosen with equal probability from among the eligible digits. The critical repetition was embedded at a random location in the sequence, so that (1) the first member of the repetition pair occurred between sequence positions 11 and 20, and (2) all other sequence constraints remained satisfied. Each sequence was generated by a new random draw.

Figure 12.2a shows a typical sequence of thirty digits. Figure 12.2b shows the expected distribution of lags in such a sequence. A single lag of 1, 2, 3, or 4 represents the to-be-detected repetition—the signal. All the other repetitions have lag ≥ 9 and represent the noise. The distribution of noise lags is approximately exponential; it is truncated because repetition lags greater than 21 are impossible. While the actual noise distribution of lags is well defined, the *effective* noise distribution depends somewhat on how precisely, in such a rapid sequence, subjects can use their knowledge of constraints on the frames in which repeated pairs are permitted to occur (see below).

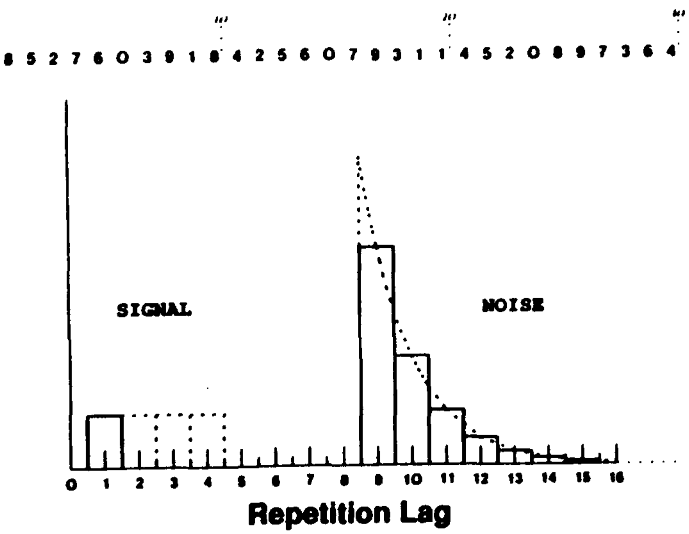


Figure 12.2 *Top*: A stimulus sequence in the repetition-detection experiment. *Bottom*: The expected frequency distribution of signal (target) and noise repetitions. Signal indicates that, on each trial, there is exactly one signal repetition; its lag is either 1, 2, 3, or 4. Nontarget digits are constrained to repeat only with lags of 9 or more (noise repetitions). The numbers 10 and 20 (top) demark the middle ten positions of the sequence within which the initial element of the target repetition is constrained to occur. These two constraints determine the expected frequency distribution of noise repetitions, indicated as NOISE.

Procedures Subjects were instructed to detect the repetition of lag 4 or less, and not to respond to any of the other stimuli. No masking stimuli were interleaved between the digits. All digits were presented in the same spatial location, centered on the CRT screen.

A trial began with a centrally located fixation square. When the subject was ready to begin the trial, the subject pressed any key on the computer keyboard. After a repetition was detected, the subject pressed the return key as quickly as possible. After the end of the sequence, a message was presented on the monitor that cued the subject to enter the repeated digit and to enter a confidence rating between 0 (very low confidence that the response was the repetition) and 4 (very high confidence that the response was the correct repetition). The actual repeated digit was then presented on the screen to give the subject complete accuracy feedback information. A message to press the Return key was displayed, following which, the fixation square for the next trial appeared.

Stimulus Sets

Subjects viewed all stimuli at a distance of 93 cm. The square fixation box was 2.46 × 2.46 degree visual angle. The digits 0 to 9 were used in the Times Roman font. The background of all displays and blank intervals was set at 50 cd/m^2. Unless otherwise specified, digits were white on gray, with a digit height of 0.74 degree.

Six stimulus dimensions were investigated separately in the experiments. There were two levels (feature values) for each of the six dimensions. The stimulus sets are shown in figure 12.3. The six dimensions (and the two feature values of each, *A* and *B*, respectively) were

1. *Size* (large, 0.74 degree visual angle versus small, 0.49 degree visual angle).

2. *Orientation* (slanted 45 degrees up-to-the-left versus slanted 45 degrees right).

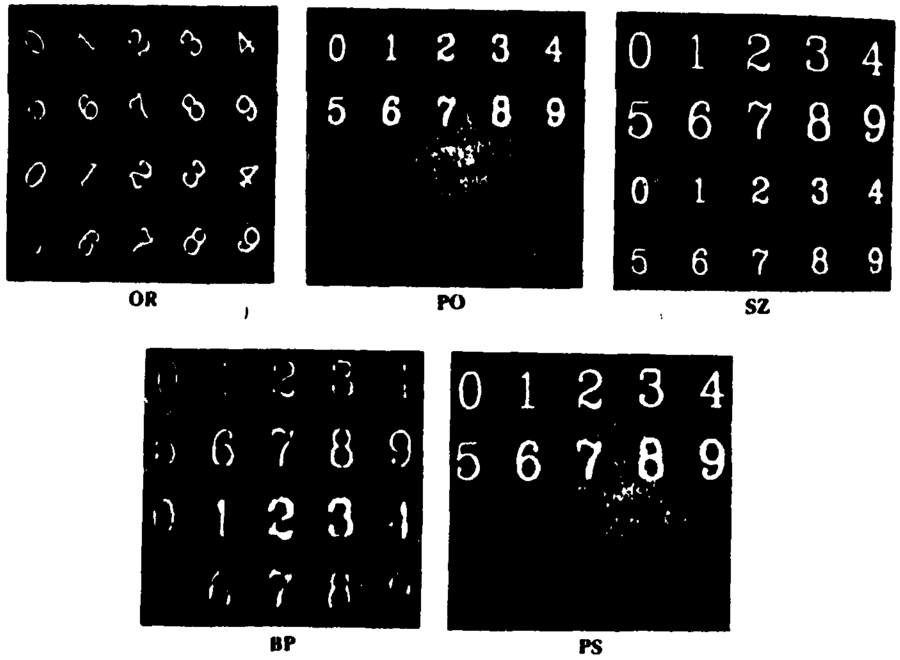


Figure 12.3 Stimuli used in the experiments. In each panel, the top ten digits are the type *A* stimulus of the indicated dimensions (orientation, polarity, size, bandpass, polarity and size). The bottom ten digits are the type *B* stimuli. Color (not shown) is similar to PO.

3. *Contrast Polarity* (white digits on gray background versus black digits on gray). The luminance level of the white digits was 101.50 cd/m^2, and the luminance level of the black digits was 0.40 cd/m^2 against a background of 50 cd/m^2.

4. *Color* (red digits on gray background vs. green digits on gray). Both red and green digits were o8 cd/m^2; saturation was chosen such that red and green were perceived as "equally different" from the background of 50 cd/m^2.

5. *Bandpass Filter* (high spatial bandpass vs. low spatial bandpass). The mean luminance level for all bandpass-filtered stimuli was 50 cd/m^2. The high bandpass digits had a mean two-dimensional frequency of 5.77 cycles per letter height, and the low bandpass digits had a frequency of 2.92 cycles per letter height. (See Parish and Sperling 1991 for a description of the filters.)

6. *Polarity and Size* Large white digits represented feature type *A*; small black digits were type *B*. All were presented against the gray background. (Large, small, light, dark, gray were as defined above.)

Blocks of Trials

Figure 12.4 illustrates the design of experimental and control stimulus sequences and presents examples. A block of trials contained only one of the six stimulus transformations (fig. 12.3). The experimental blocks all were of type $(\frac{1}{2}A + \frac{1}{2}B)$ in which streams of strictly alternating *A*, *B* stimulus features were presented. There were three kinds of experimental blocks for a given transformation that differed in the attentional conditions: attend to *A*, equal attention, attend to *B*). In addition to experimental blocks, which consisted of sequences that alternated two feature values (*A* and *B*), there were control blocks, which consisted of digits having the same feature value throughout.

Experimental blocks contained 100 trials, and control blocks contained 150 trials. Every subject ran at least four blocks in every condition (2400 trials per transformation). Each of the trials was classified according to lag 1, 2, 3, or 4. In the experimental $(\frac{1}{2}A + \frac{1}{2}B)$ blocks, trials were classified according to whether the repetition pair was *aa*, *ab*, *ba*, or *bb*. We use *A* and *B* to denote features or streams that contain the features (e.g., *A* = large and *B* = small). We use *a*, *b*, respectively, to denote target digits—members of the repetition pair—that contain features *A* and *B*, respectively.

Attention Conditions

The three experimental blocks are distinguished by the attentional instructions, the probability of the different types of repetitions presented, and the payoffs for correct responses. For the Attend-*A* experimental block the subject was instructed to devote 80 percent of attention to feature *A* (e.g., large) and 20 percent to feature *B* (e.g., small); for the Attend-*B* experimental block, the subject was instructed to devote 80 percent of attention to feature *B* (e.g., small) and 20 percent to feature *A* (e.g., large). In equal-attention experimental

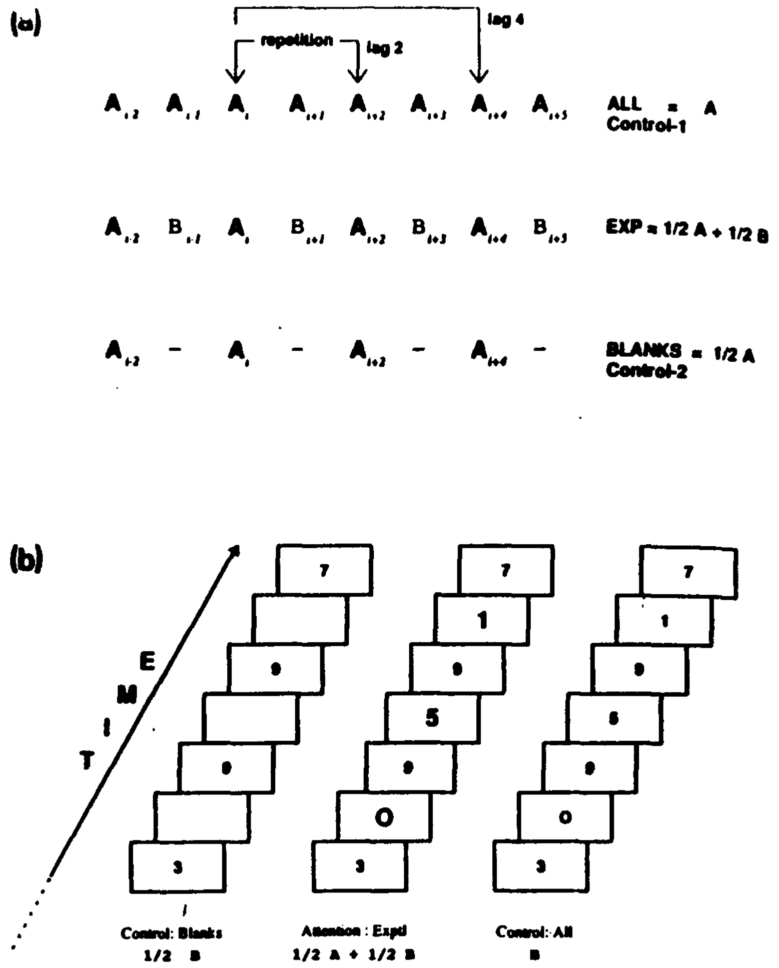

(a)
repetition
lag 2
lag 4
A_{i-2} A_{i-1} A_i A_{i+1} A_{i+2} A_{i+3} A_{i+4} A_{i+5}
ALL = A
Control-1
A_{i-2} B_{i-1} A_i B_{i+1} A_{i+2} B_{i+3} A_{i+4} B_{i+5}
EXP = 1/2 A + 1/2 B
A_{i-2} – A_i – A_{i+2} – A_{i+4} –
BLANKS = 1/2 A
Control-2
(b)
T
I
M
E
7
9
9
3
7
1
9
5
9
O
3
7
1
9
5
9
O
3
Control: Blanks
1/2 B
Attention : Exptl
1/2 A + 1/2 B
Control: All
B

Figure 12.4 Experimental and control presentation sequences used to estimate the effectiveness of attentional filtering. (a) The middle row indicates the experimental condition, an alternating sequence of type *A* and type *B* stimuli, designated as ½*A* + ½*B*. If the subject could *not* discriminate the features that distinguished the type *A* and type *B* stimuli, the subject would perform equivalently in the ½*A* + ½*B* and in the All control, which consists entirely of *A* stimuli, designated simply as *A*. On the other hand, if the subject were able to perfectly ignore the unattended-*B* feature in the ½*A* + ½*B* stream, experimental performance would be equivalent to the Blanks control, designated as ½*A*. This would be true for repetitions at lag 2 and at lag 4 (indicated above). (b) Graphical illustration of the three types of displays. The dimension is size. Type *A* stimuli are large, type *B* are small; the example illustrates *bb* targets.

Table 12.1 Probability of Each Condition within Each Block of Trials

Blocks of Alternating-Feature Sequences (*AB*)

	Attend-*A*		Attend-*B*		Attend-equal	
Target =	A	B	A	B	A	B
Lag 1*	.05	.05	.05	.05	.07	.07
Lag 2	.35	.05	.05	.35	.18	.18
Lag 3*	.05	.05	.05	.05	.07	.07
Lag 4	.35	.05	.05	.35	.18	.18

Blocks of Single-Feature Sequences

	Feature *A*		Feature *B*	
Stim. =	AA	A–	BB	B–
Lag 1	.167	—	.167	—
Lag 2	.167	.167	.167	.167
Lag 3	.167	—	.167	—
Lag 4	.167	.167	.167	.167

*Mixed-feature repetition pairs; "Target" indicates the feature of the first element of the pair.

blocks, the subject was instructed to devote 50 percent of attention to feature *A* and 50 percent to feature *B*. The probabilities of different trial types for the Attend-*A*, Attend-*B*, and Attend-equal blocks are shown in table 12.1. Note that when attending to feature *A*, 70 percent of the trials in the selective attention blocks are pure (*aa*) repetitions (35 percent at lag 2, 35 percent at lag 4). The remaining trials consist of mixed repetitions at lags 1 and 3, (*ab*) 10 percent, (*ba*) 10 percent, and of pure unattended-feature repetitions at lags 2 and 4, (*bb*) 10 percent. The converse holds when attending to feature *B*.

The attention instructions served only to define the initial conditions for the subjects. The steady-state behavior of subjects was controlled by carefully defined rewards to enforce the attention conditions. For every stimulus repetition in the attended-to stream that the subject detected correctly (that is, an *aa* or *bb* pair), the subject received five points. The subject received only one point for detecting repetitions in the unattended stream, and zero points for for the heterogeneous *ab* and *ba* repetitions. The paid subjects were paid 1 cent per point (in addition to their usual hourly wage for participation). The maximum expected payoff per trial for detecting targets with the attended feature is their probability of occurrence (0.7, table 12.1) times their value (5 cents), a net of 3.5 cents. The maximum expected earnings from detecting targets with the unattended feature is 0.1 × 1 cent = 0.1 cent. Thus, the expected value of detecting repetitions with the attended-to feature was thirty-five times greater than the value of unattended-feature repetitions. The 35 : 1 attended/unattended ratio of maximum possible earnings exerted a potent control over attention, although some of the effects of attention were unanticipated.

100 Percent–0 Percent Attention Conditions Even the extreme divided attention conditions (nominally 80 percent–20 percent) involve divided attention because, when the subject notices repetitions involving the unattended feature, they are reported. Why not include experimental conditions in which the subjects are told to give 100 percent (rather than 80 percent) of their attention to the attended feature, and to give 0 percent (rather than 20 percent) to the unattended feature, and are paid only for detecting attended-feature repetitions? In previous research, Sperling and Melchner (1978a, 1978b) compared 100 percent–0 percent attention to a range of divided attention conditions similar to the nominal 80 percent–20 percent range used here. Sperling and Melchner's attentional manipulation involved only instructions; in contrast to the present study, their instructions were unenhanced by differential probabilities of occurrence of or by differential rewards for detecting attended targets. Nevertheless, in one-third of their cases, Sperling and Melchner's (1978b) divided-attention conditions spanned a range of performances that was fully as great as the extremes of the 100 percent–0 percent control conditions, and their remaining divided-attention cases spanned most of the 100 percent–0 percent performance range. Thus, while 100 percent–0 percent conditions might (or might not) slightly expand the range of performances observed here, the added conditions would not be expected to produce any qualitatively different data.

Controls (*A*, *B*, ½*A*, ½*B*) Control blocks were run for each feature, as indicated in figure 12.4 and in table 12.1. In the control-All trials (*A* and *B*), all thirty digits have the same feature value, and lags 1, 2, 3, and 4 occur equally often. Control-All trials were interleaved with control-Blanks trials (½*A* and ½*B*) in which every other digit in the sequence was replaced by enough blank frames to permit the remaining digits to retain their precise temporal positions in the sequence. Therefore, for control-Blanks, only fifteen digits were presented, and repetitions only occurred at what, in the All sequence, would have been called lags 2 and 4 (since blanks occurred at lags 1 and 3). As indicated in table 12.1, the six control conditions with feature *A* (or feature *B*) had an equal probability of occurring (i.e., twenty-five trials for each condition in the control blocks).

Altogether, there were thirty-six different kinds of trials for each of the six stimulus transformations (fig. 12.3). There were twenty-four experimental conditions: 4 lags (1, 2, 3, 4) × 3 attentional instructions (80%, 50%, 20%) × two kinds of targets (*aa*, *bb* at lags 2, 4; *ab*, *ba* at lags 1, 3). And there were twelve control conditions: control-All contained 4 lags (1, 2, 3, 4) × 2 features (*A*, *B*), whereas control-Blanks contained 2 lags (2, 4) × 2 features (½*A*, ½*B*).

Apparatus

A desktop computer (an IBM-compatible AT personal computer) was used to present stimuli and collect subjects' responses. Stimuli were created with HIPS image-processing software (Landy, Cohen, and Sperling 1984a, b) and displayed using a software package (Runtime Library for Psychology Experiments 1988) designed to drive an AT-Vista Videographics Adapter that produced black-and-white and color images on a NEC Multisync-Plus monitor (with horizontal resolution of 960 dots, vertical resolution of 720 lines, and short persistence phosphors).

Subjects

Two female and three male New York University graduate students and staff with normal or corrected-to-normal vision participated in this research. Three of these subjects were paid for their participation, and two were experimenters. All subjects were well practiced on the repetition-detection procedure before the formal experiments began.

Experiment 2

In the procedure described so far, there are twenty repetitions (three occurrences of each digit) with only the target repetition having a lag of 1, 2, 3 or 4 and all the others having lags of 9 or greater. Experiment 2 was designed to investigate whether this aspect of the procedure was critical to the results. Three character sets were created.

1. Ten digits (as used in experiment 1).
2. Twenty-nine unique characters consisting of the ten digits plus nineteen letters. (The letters B, I, O, Q, S, V and Z were eliminated because of their similarity to digits or other letters.) When using this character set, only the critical item is repeated.
3. A set of ten randomly chosen characters from among the twenty-nine, with a new random selection being made on each trial. Sequences were composed as for the digit stream.

The physical characteristics of the stimuli were the same as in the white-on-gray transformation. On each trial, the character set (1, 2, 3) and the lag (1, 2, 3, 4) were chosen randomly and independently. There were six sessions of 100 trials for subject SW, twelve sessions for subject ZL.

12.3 RESULTS AND DISCUSSION

Experiment 2: Different Character Sets

Figure 12.5 shows the data of experiment 2. There is a typical drop of performance with increasing lag but absolutely no indication of any systematic difference in the results for the three different character sets. Most theories of memory would suggest that, by eliminating the noise repetitions, the twenty-nine-element set would greatly improve performance. However, this effect is insignificant. The robust invariance of the data despite variations in the

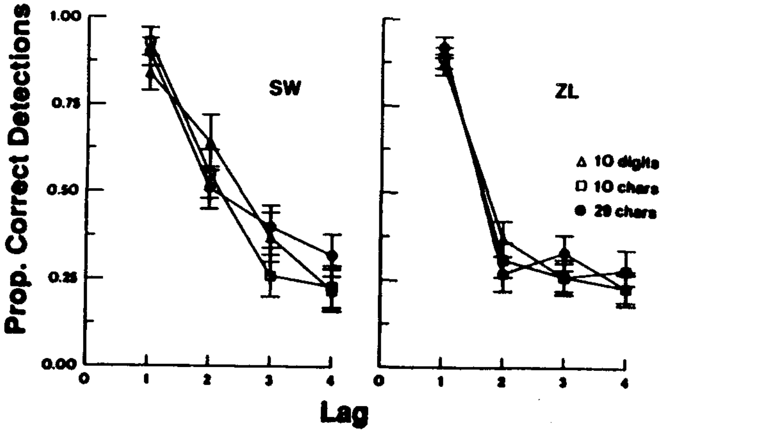


Figure 12.5 Results of variations in the character set (experiment 2). Data are shown for two subjects (SW, ZL). Lag is plotted on the abscissa, and proportion of correct detections on the ordinate. All stimulus streams contained thirty items; the curve parameter indicates the character set from which they were chosen. Open triangles—ten digits; filled circles—twenty nine characters (ten digits plus nineteen letters); open squares—a new set of ten characters chosen randomly on each trial from among the twenty nine.

nature and number of repetitions suggests that the immediate temporal environment of a repetition is the main determiner of whether or not it will be detected, and that variations in the more distant environment of a repetition are unimportant.

Experiment 1: Phenomena Illustrated with Selected Data

In the main experiment, there are thirty-six data points for each of the six types of stimuli. Therefore, presentation of the results is quite complex. We use three types of graphs. The first shows the attention conditions relative to the controls; the second shows attention-operating characteristics; and the third shows all thirty-six conditions on a single graph. We also table the benefits conferred by feature interleaving and by attentional manipulations.

Figures 12.6a, b, c show data from subject JW viewing the contrast polarity stimuli. Figure 12.6a shows detections of *aa* (white-white) repetitions in three stimulus contexts: two control stimuli ($\frac{1}{2}A$ and A, white-on-gray stimuli) and the experimental stimuli ($\frac{1}{2}A + \frac{1}{2}B$, alternating white and black stimuli). Consider first the control conditions $\frac{1}{2}A$ and A. The condition $\frac{1}{2}A$ represents a plausible upper bound on the attention conditions because it corresponds to what would be expected if the subject succeeded in ignoring B stimuli completely. The control A represents a plausible lower bound in which the B stimuli are indiscriminable from A stimuli. The projection of the diagonal line of figure 12.6a on the vertical axis (from 0.60 to 1.00) indicates the plausible bounds on the range of attention effects.

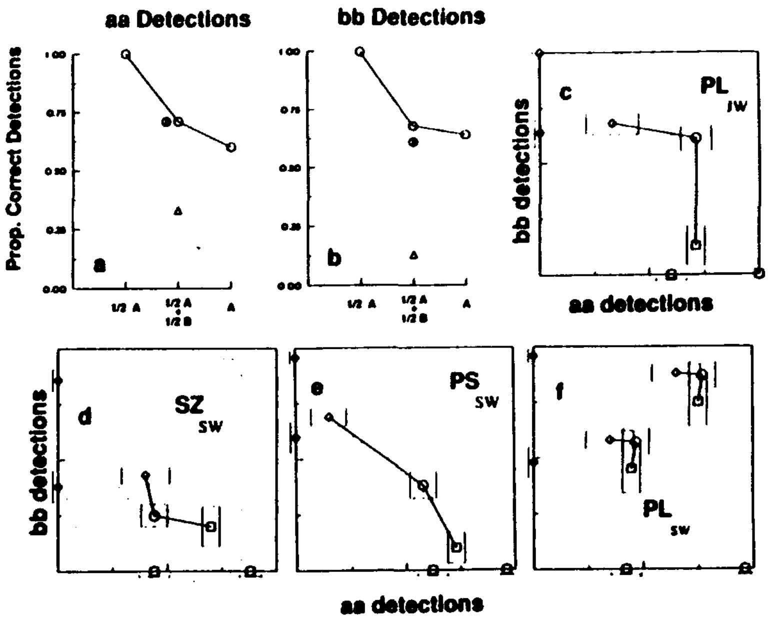


Figure 12.6 Illustrative results. (a, b, c) Polarity stimuli at lag 2 for subject JW. A = white-on-gray, B = black-on-gray. (a) The proportion correct in detecting *aa* (white-white) repetitions. Abscissa indicates the three types of stimuli (see text). The $\frac{1}{2}A + \frac{1}{2}B$ stimuli serve three attention conditions: the open circle indicates Attend-A, and it is connected by lines to the control conditions (which involve only *aa* repetitions); the half-filled point represents equal attention; the triangle indicates detecting *aa* while attending B. (b) Data for detecting type *bb* (black-black) repetitions. The open circle represents Attend-B; the half-filled point represents equal attention; the triangle indicates detecting *bb* while attending A. (c) Attention-operating characteristic (AOC) derived from the data of panels (a) and (b). The abscissa and ordinate both range from 0 to 1.0 and represent the proportions of correct *aa* and *bb* detections. The inner shaded area indicates performance worse than the corresponding *All* controls (A, B) for both *aa* and *bb* detections. The concave-down curve is the AOC derived from the $\frac{1}{2}A + \frac{1}{2}B$ stimulus with the points representing, from left-to-right, Attend-A, equal attention, and Attend-B. The error bars indicate one standard error of the mean; the relative sizes of the errors derive from the inverse square root of the number of observations. The concave-down curve of this AOC corresponds to costs but no benefits from selective attention. (d, e, f) Examples of different types of performance. (d) AOC for subject SW at lag 2 with size stimuli (A = large; type B = small). The concave-up shape of the AOC indicates benefits from selective attention without costs. The outer shaded area indicates performance better than a *Blanks* control ($\frac{1}{2}A + \frac{1}{2}B$) for one or both of the two types of targets (*aa*, *bb*). (e) AOC for subject SW at lag 2 with polarity-and-size stimuli (A = large-white, B = small-black). The AOC with slope ≈ -1 indicates symmetrical trade-offs of costs and benefits of selective attention. (f) Two AOCs are plotted: the lower-left AOC is the AOC for subject SW at lag 2 with polarity stimuli (A = white, B = black). This real AOC is "enhanced" by adding 0.3 to each coordinate to produce the "enhanced" AOC at the upper right. The real AOC indicates a small stimulus differentiation benefit; the "enhanced" AOC indicates a large benefit; attention effects are identical for both AOCs.

In the experimental conditions, $\frac{1}{2}A + \frac{1}{2}B$, full attention to feature A while ignoring B is represented by the middle point on the diagonal line of figure 12.6a. Full attention to A shows a benefit relative to the control-All-A condition but not nearly as great a benefit as occurs when the B stimuli are replaced with blanks.

Two unconnected data points are shown in figure 12.6a. The half-shaded point adjacent to the full-attention point in figure 12.6a indicates equal attention. Equal attention in an alternating $\frac{1}{2}A + \frac{1}{2}B$ stream yields better performance than in the All-A stream because mixing two features in the stream (instead of only one) makes the stimuli more discriminable. Attention to B stimuli leads to poor performance on *aa* repetitions (0.25), and this is indicated by the triangle in figure 12.6a.

We expect good symmetry between features A and B (white-on-gray, black-on-gray). Indeed, figure 12.6b, generated for detections of *bb* repetitions, is basically similar to figure 12.6a.

Generating Attention-Operating Characteristics (AOCs) The $\frac{1}{2}A + \frac{1}{2}B$ points in figures 12.6a and 12.6b generate the AOC (Kinchla 1980; Sperling and Melchner 1978b) of figure 12.6c. The lower-right square of figure 12.6c indicates joint performance on *aa* and *bb* repetitions when attention is directed to A. The rectangle around the square indicates one standard error of the mean in each dimension. The rectangle is extended in the B dimension because, in the Attend-A condition, there are seven times more *aa* repetition trials than *bb* trials, and this increases the standard error of *bb* detections relative to *aa*. The circle in figure 12.6c indicates equal-attention performance, and the diamond at the upper-left end of the AOC indicates Attend-B performance. Based on the data of figures 12.6a and 12.6b, the shape of the AOC is concave down, the limbs forming almost a right angle. The severe concave-down shape indicates that, relative to equal attention, selective attention yields negligible benefits but significant losses.

Additionally, figure 12.6c indicates a shaded area that represents excluded performances. Regardless of the state of attention, we expect performance in $\frac{1}{2}A + \frac{1}{2}B$ to equal or exceed performance in the All-A and All-B control conditions. This constraint excludes data from the lower-left rectangle of the AOC graph.

Figure 12.6d indicates an AOC derived from subject SW viewing the *size* stimuli. Here, the AOC is concave up. It indicates that, relative to equal attention, paying selective attention to large (A) stimuli improves detection of *aa* repetitions with only an insignificant loss of detectability of *bb* repetitions. Similarly, attending to small (B) stimuli significantly improves detection of *bb* repetitions but does not significantly penalize *aa* detections. A right-angle concave-up shape of AOC indicates benefits of selective attention with no costs.

Figure 12.6d illustrates a second shaded region that was absent in figure 12.6c because of that subject's perfect performance in the control conditions. Regardless of the state of attention, we expect the subject to perform worse in any experimental $\frac{1}{2}A + \frac{1}{2}B$ condition than in the corresponding $\frac{1}{2}A$ or $\frac{1}{2}B$ control condition. This excludes data from the shaded area in the outer rim of the AOC graph.

The pure costs and pure benefits indicated by the AOCs of figures 12.6c and 12.6d are somewhat unusual. Figure 12.6e (subject SW, polarity-and-size stimuli) illustrates a more typical AOC. The slope of -1 suggests a symmetrical trade-off between the costs and benefits of selective attention. The most common interpretation of linear AOCs is that the subject can perform only one task or the other, and that the equal-attention point represents a mixture of these two strategies (Sperling and Dosher 1986; Sperling and Melchner 1978a). Such a switching strategy could cause the AOC to traverse the excluded region and influence the equal-attention point to lie within it.

AOCs for the Data

Figure 12.7 shows all the AOCs from the experiments. The twenty-eight AOCs represent six stimulus transformations, each with lags 2 and 4. There are two subjects for each of the first four conditions and three subjects for the two remaining. Overall, the AOCs look similar to those illustrated in figure 12.6. Performance is consistently better for lag 2 than for lag 4.

Several AOCs show pure costs: for example, polarity (SW, JW, lag 2), and size (JW, lag 2), and many AOCs have a purely vertical or horizontal leg to indicate that one of the two selective-attention conditions results in pure costs. All in all, there are very few examples of AOCs that can be interpreted as yielding a continuum of trade-offs. We defer further discussion of these graphs until we consider the full range of data and additional summary statistics.

Consolidated Graphs of All Experimental and Control Conditions

Each panel of figure 12.8 shows mean data for each of the thirty-six kinds of repetition detections for one subject and one set of features. Except for variances and tests of significance, these graphs represent the entire data of the experiments. The plan of figure 12.8 is to indicate the data of control conditions by two sets of connected lines that form upper and lower reference bounds for four clusters of points that represent the data of the experimental conditions. We begin by making some general observations.

The Effects of Lag and SOA The effects of lag on repetition-detection performance are indicated in figure 12.8 by the sloping connecting lines that indicate performance in the control-All-A and All-B conditions. Performance with the control-All stimuli is at or above 75 percent at lag 1 for nearly all subjects and types of stimulus transformation. There are clear individual differences. For example, in the polarity-and-size conditions, subject ZL performs better at lag 1 than does subject SW, although SW had much more practice. These lag data are completely consistent with earlier observations (Kaufman 1978).

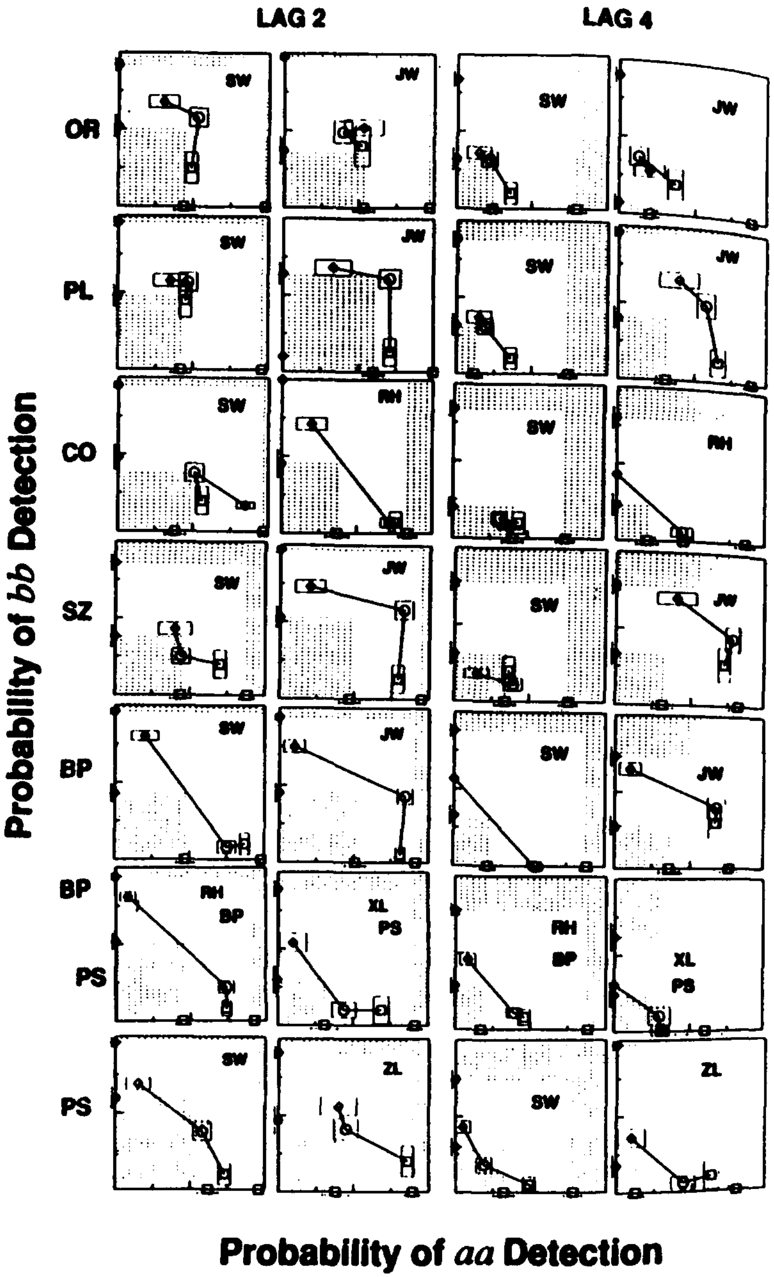


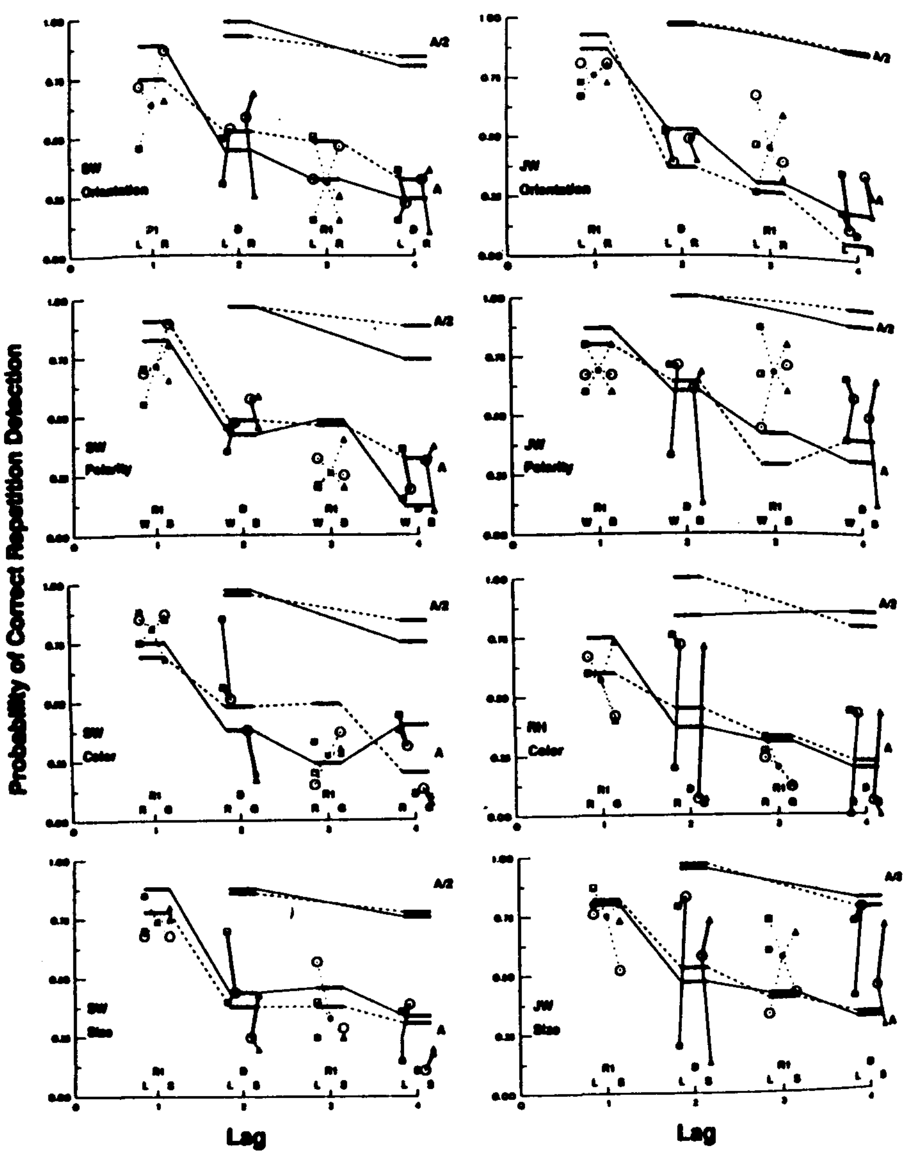
Figure 12.7 Attention-Operating Characteristics (AOC) for all subjects, stimulus transformations, and lags. Rows indicate different stimulus transformations (see figure 12.3) except row 6, which is shared by two different stimuli. The abscissa is the probability of detecting *aa* targets, the ordinate indicates *bb* detections. Coordinates range from 0 to 1.0. Symbols represent attention conditions: squares = Attend-*A*, circles = equal attention, diamonds = Attend-*B*. One standard error is indicated around each point. *A* and ½*A* control conditions are shown on the abscissa, *B* and ½*B* on the ordinate. The clear area defines the reasonable bounds on performance. SW, JW, RH, XL, and ZL indicate subjects; other abbreviations indicate transformations.

The effect of SOA is derived from the sloping lines labeled *A*/2 that represent data for the control-Blanks conditions (½*A*, ½*B*), and which appear above lags 2 and 4. Performance in control-Blanks is better than the corresponding control-All (*A*, *B*) data. Alternatively, the control-Blanks conditions with lags 2 and 4 might be described as lags 1 and 2 of a stream with a doubled SOA (stimulus onset asynchrony—the time from the onset of one digit to the next). However, the control-Blanks is not quite equivalent to a slower sequence because it has only fifteen instead of the thirty items that would be produced by simply slowing the stream. The combined manipulation of slowing and shortening the sequence produces (except for ceiling effects) better performance for the control-Blanks than the comparable control-All conditions: control-Blanks, lag 2, surpasses control-All, lag 1, and control-Blanks, lag 4, surpasses control-All, lag 2.

The obvious interpretation of these data is that the main cause of the decline of performance with lag is retroactive interference (versus passive decay). Increasing the SOA increases the amount of time that the items must be retained but actually improves performance. (We know this also from unpublished observations in our laboratory in which sequence length was controlled.)

Repetition Blindness The improvement of detection with shorter lags is different from another phenomenon discovered recently by using superficially similar procedures. "Repetition blindness" (Kanwisher 1987) is the reduced ability of subjects to report both occurrences of a repeated word embedded in a rapid sequence (approximately 4 to 9 per second) relative to the reportability of two independent words. In contrast to the present research, reportability of both occurrences of the word increases with increasing lag. There are several differences between our repetition-detection procedure and the procedure Kanwisher used. Repeated items are discriminated from unrepeated items in Kanwisher's studies rather than from other equally-often-repeated items, as in ours. However, the equivalence of the twenty-nine-element character set of experiment 2 (in which all noncritical repetitions were eliminated) to the other character sets shows that multiple repetitions are not the cause of the difference in results.

The repetition-blindness paradigm tests the tendency of subjects to report both occurrences of repeated items rather than their ability to discriminate repeated from unrepeated items. Moreover, repetition-blindness experiments typically have used linguistic stimuli (words) in the stimulus sequence, in some instances varying the context in which these words were presented (Kanwisher 1987; Kanwisher and Potter 1989), and in other in-



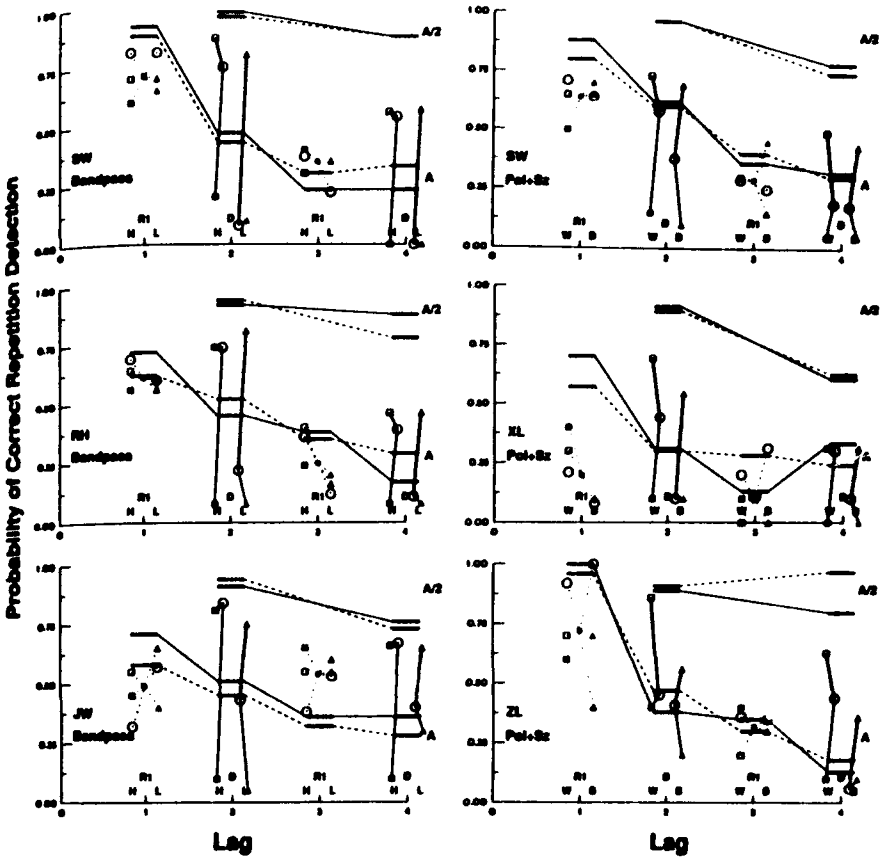


Figure 12.8 Data of all thirty-six trial types for each subject and type of stimulus transformation. In each panel, frame lag is plotted on the abscissa, and proportion of correct detections is plotted on the ordinate. Horizontal bars connected by continuous lines labeled A/2, A, represent control conditions ½*A* (control-Blanks) and *A* (control-All). ½*B* and *B* conditions are indicated by bars connected by dashed lines (not labeled). The data points at each of the frame lags represent the different attention conditions and targets in ½*A* + ½*B* stimuli. Frame lags 2 and 4 indicate *aa* and *bb* detections; frame lags 1 and 3 indicate heterogeneous *ab* and *ba* detections. Open circles indicate equal attention. At frame lags 2 and 4, data points for the detection of *aa* repetitions are displaced to the left and detections of *bb* to the right, as indicated by dimension labels below (D indicates "detection"). At frame lags 1 and 3, detections of *ab* repetitions are displaced to the left and detections of *ba* repetitions to the right. R1 indicates the first occurring feature in a heterogeneous-repetition pair, indicated by the dimension label below R1. Open symbols indicate detection of the attended feature (even lags) or detection of heterogeneous-repetition pairs in which the attended feature occurred first. Filled symbols indicate reports of unattended features or, in heterogeneous pairs, that the attended feature occurred second. Reports of *aa* (*bb*) under different attention conditions are linked by lines; the heavier line indicates the attended feature. The asterisks at frame lags 1 and 3 indicate the means for the six heterogeneous-repetition types.

stances varying the case of the repetition without incurring a performance detriment (Marohn and Hochaus 1988). These procedural and stimulus differences suggest that *repetition-blindness* and *repetition-detection* paradigms may elicit different information-processing strategies and may reflect different levels of processing.

Equivalence of the Opposed Features within a Dimension A glimpse at the control data in figure 12.8 shows that performance on the A and B control streams is essentially equivalent in all conditions. None of the differences approaches statistical significance.

Feature equivalence means that differential attentional effects exhibited in the $\frac{1}{2}A + \frac{1}{2}B$ conditions are due to factors other than differential discriminability of the streams. Further, we note that attentional effects cannot be due to cross-stream masking in which an item from $\frac{1}{2}A$ masks one from $\frac{1}{2}B$. We refer again to an earlier result that interposing noise fields between successive frames has minimal effects on performance (Kaufman 1978; Wurst, Sperling, and Dosher 1991).

Dominance Relations of Opposed Features within a Dimension Whereas the opposed A and B features are equivalent when viewed in pure stimulus streams, when they are interleaved in a $\frac{1}{2}A + \frac{1}{2}B$ stream, in the unsymmetric dimensions, one feature may dominate completely. For example, in the color dimension, red is dominant over green. When subject RH attempts to pay equal attention to both colors, she performs exactly as she does when paying selective attention to red (figure 12.7). For subject SW, the color transformation is even more problematical. He is able to selectively attend to red. However, when he attempts to selectively attend to green, his performance on green deteriorates and his performance on red improves. This result was so unexpected that extra sessions were conducted. But the additional trials merely produced more of the same kind of data.

Other examples of dominance are high bandpass over low (subjects SW, RH) and large-white over small-black (subject ZL). The dominance of one feature over another is quite similar to the dominance of one eye over another: Alone, each eye or feature may be equivalent; dominance is observed only when they are placed into competition.

Heterogeneous Detections, *ab* and *ba* In figure 12.8, heterogeneous detections are represented as clusters of points that lie above lags 1 and 3. Because of the strict feature alternation in the $\frac{1}{2}A + \frac{1}{2}B$ stream, different-feature (heterogeneous) repetitions can occur only at lags 1 and 3. The probabilities of these repetitions were quite low, $P = 0.1$ in the selective-attention conditions and $P = 0.14$ in the equal-attention condition (table 12.1). At each of these lags, there are six heterogeneous-detection types: three attentional states × two feature sequences (*ab*, *ba*). All six detection types are illustrated in figure 12.8 for each stimulus transformation, subject, and lag.

Because a heterogeneous repetition involves a feature difference, we expect heterogeneous repetitions to be more poorly detected than same-feature repetitions in all conditions (e.g., Posner et al. 1969, name vs. physical identity matching). The mean of all six heterogeneous-repetition types for lag 1 is below the level of same-feature repetitions in most instances, and dramatically below the same-feature level in some instances (with the exception of subject JW). Further characterization of heterogeneous-detection performance requires a more computational approach; we begin by developing some descriptive statistics of homogeneous detections.

Benefits and Costs in Homogeneous-Repetition Detections

A Computation Example The goal is to characterize selective attention in terms of the efficiency of attentional filtering relative to perfect optical filtering. However, selective attention is studied in the alternating stimulus $\frac{1}{2}A + \frac{1}{2}B$ in which two features are alternated. Feature alternation alone, independent of attention, may have some positive effect on performance relative to All-A or All-B controls. Therefore, we first consider the stimulus benefit of alternating features.

We begin with an illustrative computation on the data of figures 12.6a and b. Consider the range defined by the two control conditions A and $\frac{1}{2}A$. The bottom end of this range represents a point where the A and B stimuli cannot be discriminated and so performance on $\frac{1}{2}A + \frac{1}{2}B$ is equivalent to performance in either of the controls. The upper end of this range $\frac{1}{2}A$ represents the point where A and B are discriminated perfectly, and one of them can be ignored perfectly. Therefore, we expect to find attentional effects confined to this range. In figure 12.6a, the range within which benefits might be reasonably be expected to occur extends from .60 to 1.00, a range of 0.40. The equal-attention condition yields a fraction correct of .71, which is $(.71 - .60)/(1.00 - .60) = 0.28$. Attending selectively to A also yields a score of 0.71; obviously, there is no additional benefit of selective attention over equal attention. Thus we might conclude that, in detecting *aa* repetitions, there is a stimulus differentiation advantage in the $\frac{1}{2}A + \frac{1}{2}B$ stimuli relative to the All-A controls, but no advantage of selective attention.

The detection computations made on *aa* detections in figure 12.6a can be repeated for *bb* detections in figure 12.6b. There is a stimulus differentiation advantage of $(.61 - .64)/(1.00 - .64) = -0.08$, that is, a small cost. The attentional benefit $(.68 - .64)/(1.00 - .64) = +0.10$ also is small.

Finally, we average the *aa* and *bb* results to obtain a stimulus-differentiation benefit of .10 and a selective-attention benefit of 0.05; both of these differ insignificantly from zero by a t test. The conclusion is that, for these data, the performance differences between control and experimental stimuli are too small to reach statistical significance. Applying the same computations to the data of figure 12.6d yields an insignificant stimulus-differentiation benefit but a highly significant selective-attention benefit of 0.49.

In summary, the alternating-feature stream, $\frac{1}{2}A + \frac{1}{2}B$, confers two possible benefits: *stimulus discrimination* in equal-attention conditions and *attentional filtering* in selective-attention conditions. To estimate these benefits, it was useful to average the two types of detections (*aa*, *bb*).

Stimulus-Discrimination Benefit We define the *normalized stimulus-discrimination benefit* as the improvement in equal-attention conditions (equal attention minus control-All) compared to the maximum possible range of improvement (control Banks minus control-All). To compute the stimulus-discrimination benefit (Stim Disc Benefit), the following definitions are needed. Let $P(aa|\frac{1}{2}A + \frac{1}{2}B)_{Attn=A}$ be the probability of correct detections of *aa* repetitions given the $\frac{1}{2}A + \frac{1}{2}B$ stream with attention directed to the *A* feature. Let *A* indicate the All-*A* condition and $\frac{1}{2}A$ indicate the *A* blanks control condition. Then,

$$\textit{Stim Disc Benefit} = \frac{1}{2}\left[\frac{P(aa|\frac{1}{2}A + \frac{1}{2}B)_{Attn=AB} - P(aa|A)}{P(aa|\frac{1}{2}A) - P(aa|A)}\right] + \frac{1}{2}\left[\frac{P(bb|\frac{1}{2}A + \frac{1}{2}B)_{Attn=AB} - P(bb|B)}{P(bb|\frac{1}{2}B) - P(bb|B)}\right] \qquad (1)$$

Selective-Attention Benefits and Costs Similarly, the *normalized selective-attention benefit*, abbreviated here simply to Sel Attn Benefit, is

$$\textit{Sel Attn Benefit} = \frac{1}{2}\left[\frac{P(aa|\frac{1}{2}A + \frac{1}{2}B)_{Attn=A} - P(aa|\frac{1}{2}A + \frac{1}{2}B)_{Attn=AB}}{P(aa|\frac{1}{2}A) - P(aa|A)}\right] + \frac{1}{2}\left[\frac{P(bb|\frac{1}{2}A + \frac{1}{2}B)_{Attn=B} - P(bb|\frac{1}{2}A + \frac{1}{2}B)_{Attn=AB}}{P(bb|\frac{1}{2}B) - P(bb|B)}\right] \qquad (2)$$

where $Attn = AB$ denotes the equal-attention condition.

The selective-attention cost is defined exactly like the benefit in equation 2 except that the subscripts $Attn = A$ and $Attn = B$ are interchanged.

In terms of AOCs, the stimulus-discrimination benefit describes where the equal-attention point lies relative to the two forbidden areas. For example, figure 12.6f shows the AOC derived from subject SW with the polarity stimuli and the same AOC translated up and to the right. The stimulus-differentiation benefit for the real data is 0.14; for the translated data it is .67.

In terms of AOCs, the attention benefit describes how far the arms of the AOC extend outward from the equal-attention point toward the upper and far-right boundaries. For selective-attention conditions, the stimulus and attention benefits sum. Stimulus discrimination measures the extent to which the physical attributes of the items aid in making them discriminable. The selective-attention benefit measures the efficiency of attentional filtering of the unattended items. Together these factors determine how closely attention performance in $\frac{1}{2}A + \frac{1}{2}B$ approaches control performance in $\frac{1}{2}A$ and $\frac{1}{2}B$.

Benefits and Costs in Heterogeneous-Repetition Detections

Heterogeneous-Repetition Cost An alternating stimulus $\frac{1}{2}A + \frac{1}{2}B$ facilitates detections of homogeneous repetitions *aa* and *bb* because the elements of the repetition pair share a common *A* or *B* feature, and this helps to discriminate them from all the other possible pairs, half of which differ in this feature. The benefit of the $\frac{1}{2}A + \frac{1}{2}B$ stimulus becomes a cost when a heterogeneous repetition *ab* or *ba* must be detected.

To estimate the cost of heterogeneous detections, we use a computation similar to the estimation of the homogeneous stimulus-differentiation benefit. In term of the representation in figure 12.8, we measure the distance from the center of gravity of a heterogeneous cluster (the asterisk) to the mean of lower set of curves, divided by the distance between the two sets of curves. There are two complications in locating the appropriate point on the upper curve. At lag 3, we use the average of the upper curve at lag 2 and 4. At lag 1, there is no upper curve, so we simply use 1.0.

$$\textit{Hetero Rep Cost} = \frac{\frac{1}{2}(P(ab|\frac{1}{2}A + \frac{1}{2}B) + P(ba|\frac{1}{2}A + \frac{1}{2}B)) - X}{Y - X} \qquad (3)$$

where

$$X = \tfrac{1}{2}(P(aa|A) + P(bb|B))$$

$$Y_{Lag\,3} = \tfrac{1}{2}\{(P(aa|A) + P(bb|B))_{Lag\,2} + (P(aa|A) + P(bb|B))_{Lag\,4}\}$$

and

$$Y_{Lag\,1} = 1$$

For strict comparability with the homogeneous stimulus-discrimination benefit, *P*(*ab*) and *P*(*ba*) should be computed only for equal-attention conditions. However, there was so little systematic difference in heterogeneous detections between conditions that the computation is aggregated over all attention conditions.

Heterogeneous Equal-Attention Benefit In homogeneous-repetition detections, *aa*, *bb*, equal attention was, on the whole, a cost relative to selective attention. Selective attention could filter unattended stimuli prior to STVRM, thereby simplifying the task of repetition detection. Alternatively, attention could operate at the level of memory by tagging the stimuli in STVRM as "attended" or "unattended." Insofar as attention operates prior to storage in memory, attended items are benefited, and unattended items are handicapped. If attention were to operate at the level of memory, either an attended tag or an unattended tag would benefit homogeneous detections relative to heterogeneous detections. As we shall see, there were widespread costs to misdirected attention, and these costs imply an early locus for selective attention.

On the other hand, heterogeneous detections are relatively neutral to an *early locus of attention because positively directed attention favors one mem-*

ber of the repetition pair but impairs the other, and the two effects would tend to cancel. But, if attention acted at the level of memory coding, selective attention would impair heterogeneous detections because one member of the pair would be tagged as attended and the other as unattended. Whereas selective attention either at the perceptual or the memorial level would facilitate homogeneous detections, selective attention at the memorial level would be harmful to heterogeneous detections. To quantify this effect, we measure the performance difference between heterogeneous detections with selective attention and with equal attention. This difference is normalized by the same factor as the heterogeneous repetition cost. A positive equal-attention benefit for heterogeneous detections suggests a memorial locus for attention.

Tabulation of Attention Benefits and Costs

Table 12.2 presents the values of five different costs and benefits computed individually for every subject and stimulus transformation. To determine whether an effect was statistically significantly different from zero, a *t*-test was conducted on the numerator of the expression that defines the effect (equations 1–3), and these results also were tabulated. In analyzing the data we concentrate first on the regularities in the data, keeping in mind the very considerable individual differences.

Stimulus-Discrimination Benefits Overall stimulus benefits are quite small. In figure 12.7 this was indicated by the closeness of the AOCs to the lower-left forbidden area. In figure 12.8, it is indicated by tendency of the open circles that represent equal attention to fall on or near the lower curve. Nevertheless there are exceptions: six of twenty-eight stimulus-discrimination benefits are statistically significant; these occur for size, contrast polarity, and bandpass stimuli. There are significant costs for subject SW viewing polarity-and-size.

A stimulus benefit indicates that homogeneous discriminations are facilitated by common features. A cost suggests a significant inability to simultaneously attend to the opposed features, even with equal attention. Apparently, SW can attend either to large-white or to small-black stimuli, but not to both, and therefore performance in an alternating stream suffers relative to the control. (This is further borne out by the large benefits and costs he shows with selective attention for these stimuli.) On the whole, stimulus benefits are small.

Heterogeneous-Repetition Costs Stimulus benefits in homogeneous detections imply stimulus costs in heterogeneous detections. If a common physical feature aids an *aa* detection, the feature difference should impair an *ab* detection. As noted, common features aid homogeneous detections by increasing the distance of the target repetition (which has common features) from the nontarget repetitions that have differing features. This is a relatively small benefit because half of the nontarget repetitions have similar features, and these cause the interference. On the other hand, differing features impair a

Table 12.2 Benefits and Costs Achieved in Equal-Attention and in Selective-Attention Conditions

		Homog Repetition[a]: *aa, bb*						Heterog Rep: *ab, ba*					
		Stimulus		Selective		Selective		Heterog		Equal		Attn	
Stim	Sub	Discr	Benf	Attn	Benf[b]	Attn	Cost[b]	Rep	Cost	Attn	Benf	Type	
		lag = 2	4	2	4	2	4	1	3	1	3	1/2	3/4
OR	SW	.16	−.02	.09	.17*	−.58‡	−.27†	.85‡	−.18*	1.2†	.27*	s	s
	JW	−.06	.13*	.17	.12	.04	−.05	1.4†	.26*	.80	.18	o	o
PO	SW	.14	.05	−.01	.21†	−.24	−.22*	1.1†	−.50‡	.79	.09	o	p
	JW	.10	.32†	.09	.21*	−1.1‡	−.48†	.94†	.53‡	−.19	−.27*	o	r
CO	SW	.03	−.30	−.22*	.02	.29	−.09	.07	−.01	.25	−.03	r	o
	RH	.02	.03	.63‡	.33‡	−.57‡	−.38‡	−.28	−.21	−.09	−.07	p	p
SZ	SW	−.15	−.14	.49‡	.03	−.12	−.18	−.5	−.23*	−.59	.35†	p	s
	JW	.43‡	.58‡	.13	.23*	−1.2‡	−.54†	−.32	.34†	−.84*	−.62‡	r	r
BP	SW	.06	−.06	.67‡	.53‡	−.52‡	-.42‡	1.9‡	.12	1.5*	−.08	b	p
	JW	.39†	.52‡	.29‡	.27†	−1.3‡	−.80‡	.46†	.40‡	−.17	−.24	p	p
	RH	−.05	.00	.70‡	.41‡	−.87‡	−.25‡	−.18	−.24†	.15	−.02	p	p
PS	SW	−.36*	−.28*	.65‡	.61‡	−.91‡	−.30†	1.3‡	−.19*	.31	−.06	p	p
	ZL	.00	.16	.57†	.33*	−.26	−.23	−13‡	−.02	18‡	.11	b	p
	XL	−.06	−.24	.58‡	.30*	−.28*	−.69‡	1.2‡	−.20*	−.22	.42†	p	b

* Statistically significant at: * .05; †.01; ‡.001

a. Benefits and costs averaged over *aa* and *bb* repetitions.

b. Averaged over selective attention conditions.

heterogeneous target repetition relative to the half of nontarget repetitions that have similar features. This is a large cost because it brings the nearest neighbors closer. Nineteen of twenty-eight cells show significant heterogeneous repetition costs; most of these are highly significant. At lag 1, thirteen of fourteen cells show a cost, and nine of fourteen are highly significant. These data indicate that the physical feature is represented in STVRM, and that this feature representation figures prominently in repetition detection. The color dimension is an exception: color similarity seems not to play a significant role in repetition detection.

Feature similarity is a bigger effect for lag 1 than for lag 3. This is consistent with earlier observations (Kaufman 1978) that STVRM for lag 1 seems to be more iconic (less abstract) than for lag 3.

Finally, we note four significant benefits of feature dissimilarity in heterogeneous detections. These all occur for subject JW at lag 3 and characterize all his performances at this lag. Indeed, his performance with heterogeneous repetitions surpasses that of other subjects and at lag 3 surpasses his own for homogeneous repetitions. These data differ profoundly from all our other data and require a different explanation. One possibility that occurred to us is that JW uses the same repetition-detection mechanism that is used in the Kanwisher paradigm (in which longer lags aid repetition detection). If so, making the repeated item different in some physical feature might aid it in surviving repetition blindness. Of the many subjects who have run in our paradigm, JW is the only one who exhibits this effect.

Selective-Attention Benefits All subjects show highly significant attentional benefits for bandpass and polarity-and-size stimuli, and, for each of the other transformations, at least one subject shows a significant attentional benefit. The filtering efficiency of attentional filtering in the bandpass and polarity-and-size stimuli is very high. At lag 2, five of the six cells show a benefit that is 57 to 70 percent of the benefit produced by perfect optical filtering (i.e., the $\frac{1}{2}A$ and $\frac{1}{2}B$ control stimuli).

In multilocation search paradigms, it is not clear whether features merely draw attention to a location or whether information can be filtered according to physical features. In our paradigm, the data indicate that efficient attentional filtering according to physical features occurs within a single location.

Selective-Attention Costs Twenty-six of twenty-eight cells show attention costs for unattended items; nineteen of these costs are statistically significant, and there are no significant exceptions. There is, on the whole, a high correlation between attentional benefits for attended homogeneous repetitions and costs for unattended repetitions. Indeed, if detections of unattended repetitions were not correspondingly impaired, we would have to conclude that attentional selection occurred at a later stage where both attended and unattended items were available for selection.

In spite of the overall correlation between benefits for attended repetitions and costs for unattended ones, there are some obvious exceptions. Subject SW does not have an attentional benefit at lag 2 for orientation or for polarity stimuli but shows a large cost. Subject JW shows a similar effect for polarity stimuli. These observations are consistent with right-angled, concave-down AOCs (figure 12.7) for these conditions that indicate costs without benefits for selective attention.

If unattended items are filtered to the point where detection of unattended repetitions is significantly impaired, should there not be a benefit for the attended repetitions? Finding one but not both of these effects suggests that the unattended items are absent in some contexts (detecting unattended repetitions) but present in others (interfering with detection of attended repetitions). This is one of several indications in our data that attention may operate at more than one level: at a perceptual filtering level before STVRM and at the level of coding information within STVRM itself.

Equal-Attention Benefits in Heterogeneous Detections If the state of attention were coded in STVRM, then we would expect equal-attention conditions to have an advantage in heterogeneous repetitions. On the whole, equal-attention benefits are small; only nine of twenty-eight cells show statistically significant benefits. Of these, three are negative (representing costs). Costs arise in subject JW's data, and the explanation is similar to that considered for JW's heterogeneous-repetition costs. Differentiating repeated items (in this case by the state of attention) facilitates JW's repetition detection.

Patterns of Attentional Benefits

Here we consider joint attentional benefits in detection of homogeneous and heterogeneous repetitions. Figure 12.9 illustrates the four combinations of small or large selective-attention benefits in homogeneous detections with small or large equal-attention benefits in heterogeneous detections. We con-

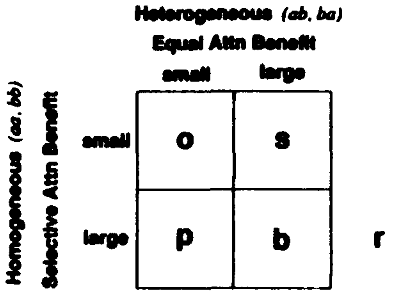


Figure 12.9 Types of attention performance, according to the joint magnitude of selective-attention benefits in homogeneous-repetition detections and equal-attention benefits in heterogeneous-repetition detections. The letters merely suggest causes: *o* = no attention benefits, *p* = perceptual benefits, *s* = benefits in STVRM (short-term visual repetition memory), *b* = both. The *r* outside the 2 × 2 table indicates reversed effects—impaired performance due to selective attention.

sider an attentional benefit to be large if it is greater than 0.20 and if it is statistically significant. Otherwise it is small. An attentional benefit that is significantly negative indicates impaired performance due to selective attention. Such effects are unexpected and are categorized separately as *r* (reversed). The last two columns of Table 12.2 use a code letter to represent the join distribution of benefits.

No Attention Benefits (*o*) There are five instances of small-homogeneous with small-heterogeneous benefits. These occur in the orientation, polarity, and color stimuli but not in any of the other conditions. The first four of these *o*'s occur in conditions in which there are very large heterogeneous-repetition costs. This demonstrates that these features are highly discriminable; the absence of an attention effect must be attributed to something else. The fifth *o* occurs for subject SW color, which we have already noted is aberrant with respect to attention: attending to green impairs SW's performance for green stimuli but improves performance for red.

Selective-Attention Benefit (*p*); Both (*b*) A selective-attention benefit implies attentional selection of attended items. Perhaps the strongest result in these experiments is the ubiquity of selective-attentional selection in certain stimulus transformations, most notably bandpass and polarity-and-size. For both of these stimuli, all three subjects, at both lags, show a strong selective-attentional benefit (twelve of twelve cells), and in three of these conditions there is also a strong equal-attention benefit (*b*). Even subjects such as JW and SW, who deal quite differently with other classes of stimuli, come together here to show strong attentional effects. Of sixteen remaining cells, only four show a *p* benefit. Clearly, the stimulus dimension strongly influences the ability of subjects to select items according to attentional instructions.

A benefit of selective attention for homogeneous detections without a corresponding penalty in heterogeneous detections (the *p* classification) suggests that attention operates prior to coding in STVRM. The *b* (both) category is ambiguous as to where attentional selection might be operating.

Equal-Attention Benefit for Heterogeneous Repetitions without a Selective-Attention Benefit for Homogeneous Repetitions (*s*) A selective-attention cost for heterogeneous pairs without a selection benefit for homogeneous pairs suggests that attentional selection is occurring in or after STVRM. (If early attentional filtering had occurred, it would have yielded a selective-attention benefit.) The three *s* cells occur when subject SW views orientation or size stimuli. These are additionally coupled with significant heterogeneous repetition cost, indicating that the physical features are represented in memory to the point of interfering with heterogeneous detections. The equal-attention benefit is, alternatively phrased, a selective-attention cost over and above the stimulus heterogeneity cost. For this subject and these stimuli, the evidence quite consistently implies that both features and the attentional state of input items are stored in STVRM.

It is noteworthy that there are significant costs of selective attention connected with two of the three *s* cells and almost significant costs in the third instance. This suggests that selective-attention costs may be occurring at the level of STVRM as well as at an earlier stage.

Reversed Effects of Selective Attention (*r*) These have already received much discussion: subject JW benefits from stimulus heterogeneity, especially at lag 3; and subject SW cannot selectively attend to green (among alternating red and green stimuli).

12.4 SUMMARY AND CONCLUSIONS

Detection of visual repetitions in a rapid stream of items depends on a short-term visual repetition memory (STVRM) that is indifferent to eye of origin and to interposed masking fields, and which functions as well for nonsense shapes as for digits. STVRM is visual, not verbal or semantic. It is governed by interference from new items; it does not suffer passive decay within the short interstimulus intervals under which it has been tested.

Using selective-attention instructions with the repetition-detection task permitted us to test the extent to which, at a single location, subjects could filter rapidly successive items according to their physical characteristics. By presenting all the items at the same location, only attentional selection according to features (and not according to location) is effective. Our subjects selectively attended to subsets of characters based on physical differences of orientation, contrast polarity, color, size, spatial bandpass filtering, and polarity-and-size combined.

Efficiency of attentional selection was determined by comparing performance in a stream of characters that alternated a physical feature with performance in two control conditions: one in which the to-be-unattended characters were optically filtered and another in which all chatacters shared the same physical feature. Selection efficiency in bandpass-filtered streams and in the polarity-and-size streams was greater than 50 percent. Attentional selection based on the other physical features was less effective or ineffective.

Corresponding to the benefits of attentional selection in detecting to-be-attended repetitions, there were large costs in the detection of unattended features. Costs were more ubiquitous than benefits.

In addition to studying repetitions of items that shared a physical feature (homogeneous repetitions), we studied heterogeneous repetitions. Costs for detecting heterogeneous repetitions (relative to homogeneous repetitions) were widespread, indicating that physical features are represented in STVRM. The corresponding stimulus benefits of detecting homogeneous repetitions in feature-alternating streams (under equal attention) were small and only occasionally significant.

If the state of attention were represented in STVRM, we would expect a cost in the detection of heterogeneous repetitions with selective-attention instructions (because the attentional state would differ for the two elements of the pair). Such costs were observed, and in some instances they occurred even when there was no corresponding benefit for selective attention in homogeneous detections. This was interpreted as a lack of early attentional filtering compensated by a memory tag representing whether or not an item was attended.

We conclude that the largest attentional effects occur at the level of attentional selection prior to encoding in STVRM (for bandpass and polarity-and-size stimuli) but that, even when early attentional filtering fails, it can still occur in STVRM.

ACKNOWLEDGMENTS

The experimental work was supported by Office of Naval Research, Cognitive and Neural Sciences Division, Grant N00014-88-K-0569; the theoretical work and preparation of the manuscript was supported by AFOSR, Life Science Directorate, Visual Information Processing Program; Grant No. 91-0178.

REFERENCES

Averbach, E., and Sperling, G. (1961). Short term storage of information in vision. In C. Cherry (Ed.), *Information Theory*, 196–211. Washington, D.C.: Butterworths.

Broadbent, D. E. (1958). *Perception and Communication*. London: Pergamon Press.

Cave, K. R., and Wolfe, J. M. (1990). Modelling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225–271.

Deutsch, J. A., and Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*, 80–90.

Duncan, J., and Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433–458.

Felfoldy, G. L., and Garner, W. R. (1971). The effects on speeded classification of implicit and explicit instructions regarding redundant dimensions. *Perception and Psychophysics, 9*, 289–292.

Folk, C. L., and Egeth, H. (1989). Does the identification of simple features require serial processing? *Journal of Experimental Psychology: Human Perception and Performance, 15*, 97–110.

Garner, W. R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General, 107*, 287–308.

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception and Psychophysics, 25*, 319–327.

Intraub, H. (1985). Visual dissociation: An illusory conjunction of pictures and forms. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 431–442.

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation *Cognition, 27*, 117–143.

Kanwisher, N. G., and Potter, M. C. (1989). Repetition blindness: The effects of stimulus modality and spatial displacement. *Memory and Cognition, 17*, 117–124.

Kaufman, J. (1978). *Visual repetition detection*. Unpublished doctoral dissertation, Department of Psychology, New York University.

Kinchla, R. A. (1980). The measurement of attention. In R. S. Nickerson (ed.), *Attention and Performance VIII*, 213–238. Hillsdale, NJ: Erlbaum.

LaBerge, D., and Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96*, 101–124.

Landy, M. S., Cohen, Y., and Sperling, G. (1984a). HIPS: Image processing under Unix. Software and applications. *Behavior Research Methods and Instrumentation, 16*, 199–216.

Landy, M. S., Cohen, Y., and Sperling, G. (1984b). HIPS: A Unix-based image processing system. *Computer Vision, Graphics, and Image Processing, 25*, 331–347.

Marohn, K. M., and Hochhaus, L. (1988). Different case repetition still leads to perceptual blindness. *Bulletin of the Psychonomic Society, 26*, 29–31.

Merikle, P. M. (1980). Selection from visual persistence by perceptual groups and category membership. *Journal of Experimental Psychology: General, 109*, 279–295.

Nakayama, K., and Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320*, 264–265.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Norman, D. A. (1968). Towards a theory of memory and attention. *Psychological Review, 75*, 522–536.

Parish, D. H., and Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31*, 1399–1410.

Pavel, M. (1991). Model of preattentive search. *Mathematical Studies in Perception and Cognition, 91–4*, New York University, Department of Psychology.

Posner, M. I., Poies, S. J., Eichelman, W., and Taylor, R. L. (1969). Retention of visual and name codes of single letters. *J. Exptl. Psychol. Monograph, 70*, 1–16.

Reeves, A., and Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93*, 180–206.

Runtime Library for Psychology Experiments. (1988). New York: HIP Lab.

Sagi, D. (1988). The combination of spatial frequency and orientation is effortlessly perceived. *Perception and Psychophysics, 43*, 601–603.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.

Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs, 74* (11, Whole No. 498).

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5*, 19–31.

Sperling, G., Budiansky, J., Spivak, J. G., and Johnson, M. C. (1971). Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science, 174*, 307–311.

Sperling, G., and Dosher, B. A. (1986). Strategy and optimization in human information processing. In K. Boff, L. Kaufman, and J. Thomas (eds.), *Handbook of Perception and Performance*. Vol. 1, 2-1–2-65. New York: Wiley.

Sperling, G., and Kaufman, J. (1978). Three kinds of visual short-term memory. Talk presented at Attention and Performance VIII, Educational Testing Service, Princeton, NJ, August 22.

Sperling, G., and Kaufman, J. (1991). Visual repetition detection. *Mathematical Studies in Perception and Cognition, 91–1*, New York University, Department of Psychology.

Sperling, G., and Melchner, M. J. (1978a). Visual search, visual attention, and the attention operating characteristic. In J. Requin (ed.), *Attention and Performance VII*, 675–686. Hillsdale, NJ Erlbaum.

Sperling, G., and Melchner, M. J. (1978b). The attention operating characteristic: Examples from visual search. *Science, 202*, 315–318.

Sperling, G., and Reeves, A. (1980). Measuring the reaction time of a shift of visual attention. In R. Nickerson (ed.), *Attention and Performance VIII*, 347–360. Hillsdale, NJ: Erlbaum.

Sperling, G., and Weichselgartner, E. (1993). Episodic theory of the dynamics of spatial attention. *Psychological Review*. In press.

Swets, J. (1984). In R. Parasuraman and D. R. Davies (eds.), *Varieties of Attention*, 183–242. New York: Academic Press.

Treisman, A. M. (1977). Focused attention in the perception and retrieval of multidimensional stimuli. *Perception and Psychophysics, 22*, 1–11.

Treisman, A. M. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 194–214.

Treisman, A. M. (1986). Properties, parts, and objects. In K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), *Handbook of Perception and Human Performance, Vol. 2*, New York: Wiley.

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

von Wright, J. M. (1968). Selection in visual immediate memory. *Quarterly Journal of Experimental Psychology, 20*, 62–68.

Weichselgartner, E. (1984). Two processes in visual attention. Unpublished doctoral dissertation. Department of Psychology, New York University.

Weichselgartner, E., and Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science, 238*, 778–780.

Wright, C. E., and Main, A. M. (1991). Selective search for conjunctively defined visual targets Unpublished manuscript.

Wurst, S. A. (1989). Investigations of short-term visual repetition memory. Unpublished doctoral dissertation, Department of Psychology, New York University.

Wurst, S. A., Sperling, G., and Dosher, B. A. (1991). The locus and process of visual repetition detection. *Mathematical Studies in Perception and Cognition, 91–2*, New York University, Department of Psychology.

13 The Skill of Attention Control: Acquisition and Execution of Attention Strategies

Daniel Gopher

13.1 PROBLEM AND SCOPE

How able and efficient are humans in controlling and utilizing their limited mental resources? Can they be taught to improve their attention management skills? Can attention control be treated as a basic skill component? These are the main questions that will be examined in the present chapter. To support the notion of attention management as a trainable skill, we need evidence to show (1) the existence of behavioral potential, namely, control over the allocation of attention, (2) difficulties and failure in fulfilling this potential, and (3) ability to overcome or diminish these difficulties with proper training.

We seek data that bear upon the ability of performers to cope efficiently with tasks that require them to divide attention and processing efforts among multiple, dynamically varying elements. Such requirements are common in many daily tasks of intermediate and high complexity, where multiple dynamic elements have to be monitored, responded to, and interacted with to achieve a common general goal. For example, car drivers are required to divide attention among the manual control of the vehicle, monitoring the road, the traffic signs, and the behavior of other vehicles, while also attempting to locate themselves geographically in the environment. A similar task with much-elevated demands is that of a pilot who moves a: high speed through six degrees of freedom space. Other examples include the task of a process controller in the control room of a modern power plant and a ba:ketball player running with the ball while looking for the best opportunity to attempt a shot or pass, while also trying to anticipate defensive moves.

Common to all these tasks is that the performer would gain most if he or she could fully attend and respond to all elements at all times. However, such full attention is not possible. Hence, some trade-offs and priorities must be established along with attention-allocation strategies. Setting priorities is a common human experience; the question is how competent we are in establishing attention strategies and allocating processing efforts among concurrently changing task elements.

A *strategy* in this context is defined as a vector of differential weights or attention biases assigned to task elements. It influences the performer's mode

Reprinted from JOURNAL OF MATHEMATICAL PSYCHOLOGY
All Rights Reserved by Academic Press, New York and London

Vol 35, No 4, December 1991
Printed in Belgium

Texture Quilts: Basic Tools for Studying Motion-from-Texture

CHARLES CHUBB

Department of Psychology, Rutgers University

AND

GEORGE SPERLING

Psychology Department and Center for Neural Sciences, New York University

A theoretical foundation and concrete stimulus-construction methods are provided for studying motion-from-spatial-texture without contamination by motion mechanisms sensitive to other aspects of the signal. Specifically, examples are constructed of a special class of random stimuli called *texture quilts*. Although, as we demonstrate experimentally, certain texture quilts display consistent apparent motion, it is proven that their motion content (a) is unavailable to standard motion analysis (such as might be accomplished by an Adelson/Bergen motion-energy analyzer, a Watson/Ahumada motion sensor, or by any elaborated Reichardt detector), and (b) cannot be exposed to standard motion analysis by any purely temporal signal transformation no matter how nonlinear (e.g., temporal differentiation followed by rectification). Applying such a purely temporal transformation to any texture quilt produces a spatiotemporal function P whose motion is unavailable to standard motion analysis: The expected response of every Reichardt detector to P is 0 at every instant in time. The simplest mechanism sufficient to sense the motion exhibited by texture quilts consists of three successive stages: (i) a purely spatial linear filter, (ii) a rectifier (but not a perfect square law) to transform regions of large negative or positive responses into regions of high positive values, and (iii) standard motion analysis. © 1991 Academic Press, Inc.

1. INTRODUCTION

Standard Motion Analysis. The extensive literature on the motion of random-dot cinematograms (Anstis, 1970; Baker & Braddick, 1982a, 1982b; Bell & Lappin, 1979; Braddick, 1973, 1974; Chang & Julesz, 1983a, 1983b, 1985; van Doorn & Koenderink, 1984; Julesz, 1971; Lappin & Bell, 1972; Nakayama & Silverman, 1984; Ramachandran & Anstis, 1983) points toward the view that a "short-range" system (Braddick, 1973, 1974) submits the raw spatiotemporal luminance function directly to *standard motion analysis* (such as might be accomplished by an Adelson/Bergen motion-energy detector (Adelson & Bergen, 1985), a Watson/Ahumada

Reprint requests should be sent to Charles Chubb, Department of Psychology, Rutgers University, New Brunswick, NJ 08903 or George Sperling, HIP Lab, NYU, 6 Washington Place, New York, NY 10003.

0022-2496/91 $3.00

motion sensor (Watson & Ahumada, 1983a, 1983b, 1985), an elaborated Reichardt detector (van Santen & Sperling, 1984, 1985), or some variants of a gradient detector (Marr & Ullman, 1981; Adelson & Bergen, 1986).

Fourier and Non-Fourier Mechanisms. An impressive number of observations suggest that standard motion analysis is not the whole story (Bowne, McKee, & Glaser, 1989; Cavanagh, Arguin, & von Grunau, 1989; Derrington & Badcock, 1985; Derrington & Henning, 1987; Green, 1986; Lelkins & Koenderink, 1984; Pantle & Turano, 1986; Petersik, Hicks, & Pantle, 1978; Ramachandran, Ginsburg, & Anstis, 1983; Ramachandran, Rao, & Vidyasagar, 1973; Sperling, 1976; Turano & Pantle, 1989). In particular, Chubb and Sperling (1987, 1988) have demonstrated a variety of stimuli that display consistent, unambiguous apparent motion, yet that do not systematically stimulate mechanisms that apply standard motion analysis directly to luminance. For reasons that become clear in Section 2, we call any motion system that applies standard analysis to the raw signal as a *Fourier* mechanism, and we refer to any system that applies standard analysis to a nonlinear transformation of the signal as a *non-Fourier* mechanism.

Microbalanced Stimuli. The methods used by Chubb & Sperling to construct stimuli whose obvious and consistent motion content cannot be revealed by applying standard motion analysis directly to luminance are founded on the notion of a *microbalanced* random stimulus. In Section 2.3.5, we show that the expected response of any standard motion analyzer applied directly to any microbalanced random stimulus is equal to the expected response of the corresponding analyzer tuned to motion of the same type, but in the opposite direction.

Microbalanced random stimuli allow us to differentially stimulate non-Fourier motion mechanisms without systematically engaging Fourier mechanisms. This is the source of their importance in the study of motion perception.

There are probably several types of non-Fourier motion mechanisms, distinguished by the different transformations they apply to the signal prior to standard motion analysis. In this paper, we extend the theory of microbalanced random stimuli in order to develop methods for constructing stimuli that selectively engage specific classes of non-Fourier mechanisms without stimulating either Fourier mechanisms or other classes of non-Fourier mechanisms.

Pointwise Transformations; Static Nonlinearities. A transformation T is called *pointwise* if the output of T at any point (x, y, t) in space-time depends only on the (stimulus) input value at that point. A *nonlinear* pointwise transformation sometimes is called a *static nonlinearity*. For instance, simple rectifiers and thresholders are pointwise transformations. In Section 3, we address the problem of isolating the class of non-Fourier mechanisms that apply a simple pointwise transformation prior to standard motion analysis from the class of all those mechanisms that apply more complicated transformations. The central result in this section is proposition 3.2 which provides necessary and sufficient conditions for a random stimulus I to be such that any pointwise transformation of I is microbalanced.

Purely Temporal Transformations and Texture Quilts. The results with pointwise transformations are extended in Section 4 to purely temporal transformations (defined in Section 2.2). Whereas, for a pointwise transformation, the transformed value at the point (x, y, t) depends only on the stimulus value at (x, y, t), in a purely temporal transformation the transformed value at (x, y, t) may depend in any way whatsoever on the entire history of stimulus values at (x, y). We define the class of stimuli called *texture quilts* (Definition 4.1) whose importance derives from the fact (proven in proposition 4.3) that any purely temporal transformation of a texture quilt is microbalanced. Concrete methods are provided for constructing *binary* and *sinusoidal* texture quilts that display consistent motion.

In Section 5, these construction methods are applied in an experiment designed to demonstrate the effectiveness of three textural properties as carriers of motion information. The textural properties are (i) spatial frequency variation, (ii) orientation variation, and (iii) variation between perceptually distinct textures with identical expected energy spectra.

2. Preliminaries

This section states the background facts presupposed by the main discussion of the paper.

2.1. *Discrete Dynamic Visual Stimuli*

Notation. Let $\mathbf{R}$ denote the real numbers, and $\mathbf{Z}$ ($\mathbf{Z}^+$) the integers (positive integers). We use square brackets to enclose arguments of discrete functions, and parentheses to enclose arguments of continuous functions.

The Range of a Stimulus. We want the term "stimulus" to refer not only to the luminance function submitted as input to the retina, but to any physiologically reasonable transformation of the spatiotemporal luminance function which might be submitted as input to a component processor of the visual system. Consequently, although luminance is physically a nonnegative quantity, we do not apply this constraint to the class of functions we admit as stimuli. We allow stimuli to take values throughout the positive and negative real numbers.

The Domain of a Stimulus. To remain close to our intuitions about neurally realized visual processors, we take stimuli to be a functions of the discrete domain $\mathbf{Z}^3$ (where the dimensions correspond to horizontal and vertical space, and time). In addition, for mathematical convenience, and without loss of physiological plausibility, we require a stimulus to be 0 almost everywhere in its (infinite) domain.

The Definition of a Stimulus. We call any function $I\colon \mathbf{Z}^3 \to \mathbf{R}$ a *stimulus* provided $I[x, y, t] = 0$ for all but finitely many points of $\mathbf{Z}^3$.

We shall be considering stimuli as functions of two spatial dimensions x, y and time t.

Stimulus Contrast. As is now well established (e.g., Shapley & Enroth-Cugell, 1984), early retinal gain-control mechanisms pass not stimulus luminance, but rather a signal approximating stimulus *contrast*, the normalized deviation at each time t of luminance at each point (x, y) in the visual field from a "background level," or "level of adaptation," which reflects the average luminance over points proximal to (x, y, t) in space and time. Because the transformation from luminance to contrast is a processing stage that is general to all of vision, we shall drop reference to mean luminance L_0, and characterize L only by its *contrast modulation function*, C:

$$C = \frac{L}{L_0} - 1. \tag{1}$$

What we argue in this paper is that the broad-band spatial filtering that mediates the step from luminance to contrast is succeeded by additional filtering stages in which a number of *narrowly tuned* spatial filters are applied to the visual signal, their output rectified, and the resulting spatiotemporal signal processed for motion information.

The History of a Stimulus at a Point in Space. For any stimulus I, any point $(x, y) \in \mathbf{Z}^2$, we define $I_{(x,y)}$, the *history of I at (x, y)*, by setting

$$I_{(x,y)}[t] = I[x, y, t] \tag{2}$$

for all $t \in \mathbf{Z}$.

Space-Time Separable Stimuli. A stimulus I is called *space-time* separable iff I can be expressed as the product of a spatial function $f\colon \mathbf{Z}^2 \to \mathbf{R}$ and a temporal function $g\colon \mathbf{Z} \to \mathbf{R}$: For all $(x, y, t) \in \mathbf{Z}^3$, $I[x, y, t] = f[x, y]\, g[t]$.

The Fourier Transform of a Stimulus. Because any stimulus I is nonzero at only a finite number of points, the energy in I is finite, implying that I has a well-defined Fourier transform.

We denote I's Fourier transform by $\hat{I}$: writing j for the complex number $(0, 1)$,

$$\hat{I}(\omega, \theta, \tau) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} I[x, y, t]\, e^{-j(\omega x + \theta y + \tau t)}. \tag{3}$$

Although $\hat{I}$ is defined for all real numbers ω, θ, τ, it is periodic over 2π in each argument. This fact is reflected in the inverse transform:

$$I[x, y, t] = \frac{1}{(2\pi)^3} \int_0^{2\pi} \int_0^{2\pi} \int_0^{2\pi} \hat{I}(\omega, \theta, \tau)\, e^{j(\omega x + \theta y + \tau t)}\, d\omega\, d\theta\, d\tau. \tag{4}$$

In the Fourier domain, we consistently use ω to index frequencies relative to x, θ frequencies relative to y, and τ frequencies relative to t.

The Function **0**. We write **0** for any function that assigns 0 to each element in its domain. Thus, **0** defined on $\mathbf{Z}^3$ is the stimulus that is zero throughout space and time. We also write **0** for the temporal function that sets $\mathbf{0}[t] = 0$ for all $t \in \mathbf{Z}$.

2.2. *Mappings and Stimulus Transformations*

Let Ω be the set of all real-valued functions of $\mathbf{Z}^3$, and call any function of Ω into Ω a *mapping*. (We shall need the general notion of a mapping only briefly in order to specify the subset of well-behaved mappings called transformations.) For any mapping M and any $I \in \Omega$, $M(I)$ is a real-valued function of $\mathbf{Z}^3$; accordingly, we write $M(I)[x, y, t]$ for the value of $M(I)$ at any point $(x, y, t) \in \mathbf{Z}^3$.

If it is continuous, a function $f\colon \mathbf{R} \to \mathbf{R}$ submits to a wide range of useful operations. For instance, if f is continuous, it can be integrated over any finite interval. Of course, f need not be continuous to meet this condition. For instance, f is integrable over any finite interval if f is discontinuous at only a finite number of points in any finite interval. If f is integrable over any finite interval, and if f also is bounded, then for any function g for which $\int_{\mathbf{R}} g$ converges, $\int_{\mathbf{R}} fg$ also converges. In particular, $\int_{\mathbf{R}} fg$ converges if g is a density function. For the results reported here, we restrict our attention to a special class of mappings, which we shall call stimulus transformations, that have properties analogous to those of the well-behaved function f. We specify these desirable properties in the following paragraph.

Continuous Mappings; Finitely Integrable Mappings; Bounded Mappings. For any $I \in \Omega$, any $p \in \mathbf{R}$, any $\psi \in \mathbf{Z}^3$, we write $I_{\psi \leftarrow p}$ for the element of Ω that is identical to I at all locations of $\mathbf{Z}^3$ except ψ, where it takes the value p. Any mapping M is called *continuous* if $M(I_{\psi \leftarrow p})[\zeta]$ is a continuous function of p for any $I \in \Omega$, and any $\psi, \zeta \in \mathbf{Z}^3$. M is called *finitely integrable* if, for any such I, ψ, and ζ, $M(I_{\psi \leftarrow p})[\zeta]$ is an integrable function of p over any finite interval. Finally, M is called *bounded* if, for any such I, ψ, and ζ, $M(I_{\psi \leftarrow p})[\zeta]$ is a bounded function of p over the set of real numbers.

Definition of a Stimulus Transformation. A *stimulus transformation* (which we shall often refer to simply as a *transformation*) is a bounded, finitely integrable, mapping T such that $T(S)$ is a stimulus for any stimulus S, and $T(\mathbf{0}) = \mathbf{0}$.

There are other reasonable constraints we might impose on the notion of a stimulus transformation. For instance, we might require a stimulus transformation to be time-invariant and causal. However, we do not include these conditions in our definition because they are not required for the results we report.

Purely Temporal Stimulus Transformations. Let Ω_T be the set of all functions mapping $\mathbf{Z}$ into $\mathbf{R}$. A transformation H is called *purely temporal* iff there exists a function $H_T\colon \Omega_T \to \Omega_T$ such that for any stimulus I, any $(x, y, t) \in \mathbf{Z}^3$,

$$H(I)[x, y, t] = H_T(I_{(x,y)})[t]. \tag{5}$$

That is, the value at the point $(x, y, t) \in \mathbf{Z}^3$ that results from applying H to I depends only on the history of I at (x, y). Since it is obvious from the context, we drop the distinction between H and H_T, and allow H to be applied both to full-fledged stimuli and to simple functions of time. Thus, for any temporal function $P: \mathbf{Z} \to \mathbf{R}$, we shall write $H(P)$ to indicate the temporal function $H_T(P)$.

We shall be particularly concerned with two types of transformations: *pointwise* transformations and *linear, shift-invariant* transformations.

Pointwise Transformations and Rectifiers. For any functions $f: A \to B$ and $g: B \to C$, the *composition* $g \bullet f: A \to C$ is given by

$$g \bullet f(a) = g(f(a)) \tag{6}$$

for any $a \in A$. For any $f: \mathbf{R} \to \mathbf{R}$, we call the mapping $f\bullet$, yielding the spatiotemporal function $f \bullet I$ when applied to stimulus I, a *pointwise* mapping (because its output value at any point in space-time depends only on its input value at that point).

As is evident, $f\bullet$ is a transformation iff (i) $f(0) = 0$, (ii) f is bounded on $\mathbf{R}$, and (iii) f is integrable over any bounded real interval. A transformation $f\bullet$ is called a *positive half-wave rectifier* if f is monotonically increasing, and $f[v] = 0$ for all $v \leqslant 0$; $f\bullet$ is called a *negative half-wave rectifier* if f is monotonically decreasing, and $f[v] = 0$ for $v \geqslant 0$. Finally, $f\bullet$ is called a *full-wave rectifier* if f is a monotonically increasing function of absolute value.

Linear, Shift-Invariant (LSI) Transformations. For any offset $\psi \in \mathbf{Z}^3$, define the mapping S^ψ by

$$S^\psi(I)[\zeta] = I[\zeta - \psi] \tag{7}$$

for any $I \in \Omega$. Thus $S^\psi(I)$ is derived by shifting I by the offset ψ in $\mathbf{Z}^3$. Any mapping M is called *shift-invariant* iff

$$S^\psi(M(I)) = M(S^\psi(I)) \tag{8}$$

for any $\psi \in \mathbf{Z}^3$, any $I \in \Omega$. In addition, M is *linear* iff for any $I, J \in \Omega$, any real numbers κ and λ

$$M(\kappa I + \lambda J) = \kappa M(I) + \lambda M(J). \tag{9}$$

As is well known, any linear, shift-invariant (LSI) transformation can be expressed as a *convolution*, which is defined for any $u \in \mathbf{Z}^3$ by

$$(k * I)[u] = \sum_{v \in \mathbf{Z}^3} k[u - v]\, I[v], \tag{10}$$

for some $k: \mathbf{Z}^3 \to \mathbf{R}$. The function k is called the *impulse response* of the transformation $k*$.

2.3. *Random Stimuli*

For any real random variable X with density f, we write $E[X]$ for the *expectation of* X:

$$E[X] = \int_{\mathbf{R}} x f(x)\, dx. \tag{11}$$

The notion of a random stimulus generalizes that of a (nonrandom) stimulus in that the values assigned points in space-time by a random stimulus are random variables (with finite variances) rather than constants.

Definition of a Random Stimulus. Call any family $\{R[x, y, t] \,\|\, (x, y, t) \in \mathbf{Z}^3\}$ of jointly distributed random variables a *random stimulus* provided

(i) $R[x, y, t]$ is constant and equal to 0 for all but finitely many $(x, y, t) \in \mathbf{Z}^3$, and

(ii) $E[R[x, y, t]^2]$ exists for all $(x, y, t) \in \mathbf{Z}^3$.

As with nonrandom stimuli, we write $\tilde{R}$ for the Fourier transform of any random stimulus R; and, for any $\chi = (x, y) \in \mathbf{Z}^2$ we write R_χ for the temporal random function defined by

$$R_\chi[t] = R[\chi, t] \tag{12}$$

for all times $t \in \mathbf{Z}$.

Space-Time Separable Random Stimuli. We call a random stimulus R *space-time separable* iff R is space-time separable with probability 1.

Constant Stimuli. Any ordinary stimulus can be regarded as a random stimulus that does not vary across independent realizations. We call such unvarying stimuli *constant*.

The Motion-from-Fourier-Components Principle. *Parseval's relation* states that the energy in a stimulus is proportional to the energy in its Fourier transform. Individual spatiotemporal Fourier components are drifting sinusoidal gratings. Thus, we can add up the energy in a dynamic visual stimulus either point-by-point in space-time, or drifting sinusoid by drifting sinusoid. A commonly encountered rule of thumb (van Santen & Sperling, 1985; Watson & Ahumada, 1983b; Watson, Ahumada, & Farrell, 1986) for predicting the apparent motion of an arbitrary stimulus $I[x, y, t] = f[x, t]$ (constant in the vertical dimension of space), is the *motion-from-Fourier-components* principle: For I regarded as a linear combination of drifting sinusoidal gratings, if most of I's energy is contributed by rightward-drifting gratings, then perceived motion should be to the right. If most of the energy resides in the leftward-drifting gratings, perceived motion should be to the left. Otherwise I should manifest no dec[illegible]ion in either direction.

Drift-Balanced Random Stimuli. The class of *drift-balanced* random stimuli (Chubb & Sperling, 1987, 1988) provides a rich pool of counterexamples to the motion-from-Fourier-components principle. A random stimulus R is drift balanced iff the expected energy in R of each drifting sinusoidal component is equal to the expected energy of the component of the same spatial frequency, drifting at the same rate, but in the opposite direction. The term *drift balanced* is defined formally as follows.

DEFINITION OF A DRIFT-BALANCED RANDOM STIMULUS. Call any random stimulus R *drift balanced* iff

$$E[|\tilde{R}(\omega, \theta, \tau)|^2] = E[|\tilde{R}(\omega, \theta, -\tau)|^2] \qquad (13)$$

for all $(\omega, \theta, \tau) \in \mathbf{R}^3$.[1]

Thus, for any class of spatiotemporal linear receptors tuned to stimulus energy in a certain spatiotemporal frequency band, a drift-balanced random stimulus will, on the average, stimulate equally well those receptors tuned to the corresponding band of opposite temporal orientation.

Microbalanced Random Stimuli. Consider the following two-flash stimulus S: In flash 1, a bright spot (call it Spot 1) appears. In flash 2, Spot 1 disappears, and two new spots appear, one to the left and one symmetrically to the right of Spot 1. As one might suppose, S is drift balanced. On the other hand, it is equally clear that a Fourier motion detector whose spatial reach encompasses the location of Spot 1 and only *one* of the Spots in flash 2 may well be stimulated in a fixed direction by S. Thus, although S is drift balanced, some Fourier motion detectors may be stimulated strongly and systematically by S. These detectors can be differentially selected by *spatial windowing*, and thereby the drift-balanced stimulus S is converted into a non-drift-balanced stimulus by multiplying it by an appropriate space-time separable function. The following subclass of drift-balanced random stimuli cannot be made non-drift-balanced by space-time separable windowing.

DEFINITION OF A MICROBALANCED RANDOM STIMULUS. Call any random stimulus I *microbalanced* iff the product WI is drift balanced for any space-time separable function W.

One can think of the multiplying function W as a "window" through which a spatiotemporal subregion of I can be "viewed" in isolation. The space-time separability of W ensures that W is "transparent" with respect to the motion-content of the region to which it is applied: W does not distort I's motion with any motion content of its own. The fact that I is microbalanced means that any subregion of I encountered through a "motion-transparent window" is drift balanced.

[1] For a proof that the expected energy of the Fourier transform of any random stimulus is everywhere well defined see Chubb & Sperling (1988, Appendix A).

The following characterization of the class of microbalanced random stimuli, and all other results stated without proof in this section, are from Chubb and Sperling (1988).

2.3.1. *A random stimulus I is microbalanced if and only if*

$$E[I[x, y, t]\, I[x', y', t'] - I[x, y, t']\, I[x', y', t]] = 0 \qquad (14)$$

for all $x, y, t, x', y', t' \in \mathbf{Z}$.

Some other relevant facts about microbalanced random stimuli:

2.3.2. *For any independent microbalanced random stimuli I and J,*

I. *the product IJ is microbalanced,*

and

II. *the convolution $I * J$ is microbalanced.*

2.3.3. (a) *Any space-time separable random stimulus is microbalanced;* (b) *any constant microbalanced stimulus is space-time separable.*

The following result is useful in constructing a wide range of microbalanced random stimuli which display striking apparent motion.

2.3.4. *Let Γ be a family of pairwise independent, microbalanced random stimuli, all but at most one of which have expectation* $\mathbf{0}$. *Then any linear combination of Γ is microbalanced.*

Reichardt Detectors and Microbalanced Random Stimuli. Two Fourier motion detectors proposed for psychophysical data (Adelson & Bergen, 1985; Watson & Ahumada, 1983a, 1983b) can be recast as *Reichardt detectors* (Adelson & Bergen, 1985; van Santen & Sperling, 1985). The Reichardt detector has many useful properties as a motion detector without regard to its specific instantiation (van Santen & Sperling, 1984, 1985).

Figure 1 shows a diagram of the Reichardt detector. It consists of spatial receptors charact rized by spatial functions f_1 and f_2, temporal filters $g_1 *$ and $g_2 *$, multipliers, a differencer, and another temporal filter $h *$. The spatial receptors f_i, $i = 1, 2$, act on the input stimulus I to produce intermediate outputs,

$$y_i[t] = \sum_{(x, y) \in \mathbf{Z}^2} f_i[x, y]\, I[x, y, t]. \qquad (15)$$

At the next stage, each temporal filter $g_j *$ transforms its input y_i $(i, j = 1, 2)$, yielding four temporal output functions: $g_j * y_i$. The left and right multipliers then compute the products

$$[y_1 * g_1[t]][y_2 * g_2[t]] \quad \text{and} \quad [y_1 * g_2[t]][y_2 * g_1[t]], \qquad (16)$$

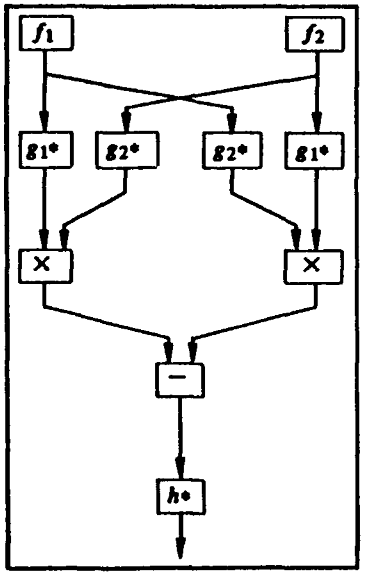


FIG. 1. The Reichardt detector. Let I be a random stimulus. Then, in response to I, for $i = 1, 2$, the box containing the *spatial function* $f_i: \mathbf{Z}^2 \to \mathbf{R}$, outputs the temporal function, $\sum_{(x,y)\in \mathbf{Z}^2} f_i[x, y]\, I[x, y, t]$; each of the boxes marked g_i* outputs the convolution of its input with the temporal function $g_i: \mathbf{Z} \to \mathbf{R}$; each of the boxes marked with a multiplication sign outputs the product of its inputs; the box marked with a minus sign outputs its left input minus its right; and the box containing $h*$ outputs the convolution of its input with the temporal function $h: \mathbf{Z} \to \mathbf{R}$. To see how the Reichardt detector senses motion, suppose f_2 is identical to f_1, but shifted in space by some offset, and suppose the filters g_1* do not alter their input, while the filters g_2* simply delay their input by some amount δ_t of time. Then a rigidly translating pattern moving in the direction of box f_2's offset from box f_1 will elicit some time-varying response from box f_1, and the same response a short time later from box f_2. If that "short time later" is precisely δ_t, the output of the righthand multiplier will be positive as long as the pattern keeps drifting. This will result in a net negative Reichardt detector output. If the pattern drift is in the opposite direction, the detector response will be positive.

respectively, and the differencer subtracts the output from the right multiplier from that of the left multiplier:

$$D[t] = [y_1 * g_1[t]][y_2 * g_2[t]] - [y_1 * g_2[t]][y_2 * g_1[t]]. \qquad (17)$$

The final output is produced by applying the filter $h*$, whose purpose is to smooth the time-varying, differencer output D. Since many Fourier mechanisms can be expressed as, or closely approximated by, Reichardt detectors (Adelson & Bergen, 1985, 1986; van Santen & Sperling, 1985), the following characterization of the class of microbalanced stimuli can be regarded as the cornerstone of the claim that microbalanced random stimuli bypass Fourier motion mechanisms.

2.3.5. *For any random stimulus I, the following conditions are equivalent:*

(I) *I is microbalanced.*

(II) *The expected response of every Reichardt detector to I is 0 at every instant in time.*

Proof. Chubb & Sperling (1988) proved that I implies II. To obtain the reverse implication, note that if II holds, then, in particular, for any points (x, y), $(x', y') \in \mathbf{Z}^2$ and any $\delta_t \in \mathbf{Z}$, the expected response to I is the temporal function **0** for a particular simple Reichardt detector that computes

$$I[x, y, t]\, I[x', y', t - \delta_t] - I[x, y, t - \delta_t]\, I[x', y', t]. \qquad (18)$$

This Reichardt detector is constructed by making (i) f_1 (of Fig. 1) the function that takes the value 1 at (x, y) and 0 everywhere else, (ii) f_2 the function that takes the value 1 at (x', y') and 0 everywhere else, (iii) each of g_1* and $h*$ the identity transformation, and (iv) g_2* the filter that delays its input by δ_t units of time. However, if the expected response to I is 0 throughout time for any such Reichardt detector, then Eq. (14) holds, and proposition 2.3.1 implies that I is microbalanced. ▮

3. Random Stimuli Microbalanced under All Pointwise Transformations

The main purpose of this paper is to provide tools for differentially stimulating specific types of non-Fourier motion mechanisms without engaging either Fourier mechanisms or other types of non-Fourier mechanisms. A non-Fourier motion mechanism is one that applies an initial nonlinear transformation to the visual signal and subjects the output to standard motion analysis. In this section, we provide some results relevant to the psychophysical problem of stimulating non-Fourier mechanisms whose initial transformation is nonpointwise without engaging any mechanism whose initial transformation is pointwise. The main finding is stated in proposition 3.2, which provides necessary and sufficient conditions for a random stimulus I to be such that $f \bullet I$ is microbalanced for any pointwise transformation $f\bullet$. In Section 4 we apply this result to construct random stimuli (texture quilts) which are microbalanced, and are, moreover, guaranteed to remain microbalanced after any purely temporal transformation. Such stimuli are useful for selectively stimulating non-Fourier motion mechanisms that extract motion information from stimuli that have undergone nonlinear *spatial* stimulus transformations.

We begin by considering an example of a stimulus (Chubb & Sperling, 1987, 1988) that is microbalanced under all pointwise transformations, but whose motion can be revealed by a purely temporal nonlinear transformation.

3.1. *Stimulus J: Traveling Reversal of a Random Black-or-White Vertical Bar Pattern.* Let $M \in \mathbf{Z}^+$. We construct the random stimulus J of $M + 1$ frames indexed 0, 1, ..., M, each of which contains M vertical bars, indexed 1, 2, ..., M from left to right. In frame 0 of stimulus J, all M vertical bars first appear. The contrast of each bar is 1 or -1 with equal probability, and bar contrasts are jointly independent. In each successive frame m, $m = 1, 2, ..., M$, the mth rectangle flips its contrast to 1 if its previous contrast was -1; otherwise it flips from 1 to -1. In frame 1, rectangle 1 flips contrast; in frame 2, rectangle 2 flips, and in successive frames, successive rectangles flip contrast from left to right, until the Mth rectangle flips in frame M, after which all the rectangles turn off. An xt cross-section of frames 0 to M of J is shown in Fig. 2a.

The traveling contrast-reversal, stimulus J, is easily expressed as a sum of pairwise independent, space-time separable random stimuli, all with expectation 0; thus propositions 2.3.3a and 2.3.4 imply that J is microbalanced. Moreover, it is easy to see that, because J's frames are comprised of only two values, any pointwise transformation of J merely serves to rescale each of J's frames, and to shift it by a constant: that is, for any $f: \mathbf{R} \to \mathbf{R}$, $f \bullet J = \lambda J + K$, where $\lambda \in \mathbf{R}$, and K is a stimulus

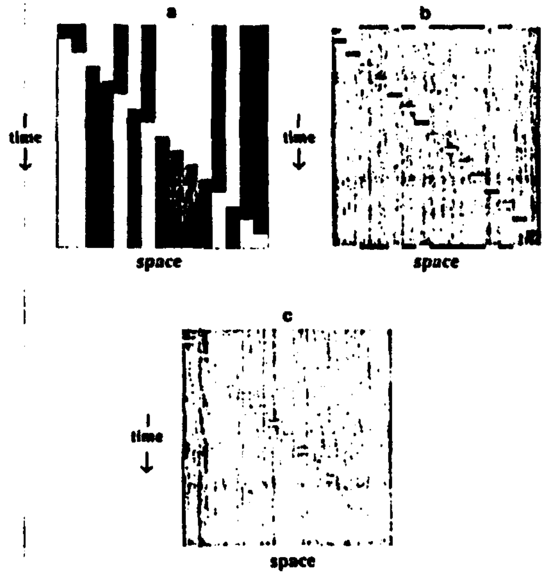


FIG. 2. Exposing the motion of the traveling contrast-reversal of the random black-or-white vertical bar pattern J to standard motion-analysis. (a) An xt cross-section of J. (b) An xt cross-section of the partial derivative of J with respect to time. (c) An xt cross-section of $|\partial J/\partial t|$. Each of J and $\partial J/\partial t$ is microbalanced. However, $|\partial J/\partial t|$ is not. In particular, $|\partial J/\partial t|$ has most of its energy at those frequencies whose velocity is equal to the velocity of the traveling contrast-reversal.

that assigns a constant value across all points at which J is nonzero. Clearly, $f \bullet J$ is another microbalanced random function (this follows easily from proposition 2.3.4). Thus, pointwise transformations fail to expose J's motion.

Exposing J's Motion to Standard Analysis. Perhaps the simplest way to extract J's motion is to full-wave rectify the partial derivative of J taken with respect to time. The stages of this transformation are illustrated in Figs. 2b and 2c. Figure 2b shows $\partial J/\partial t$. This function is itself microbalanced (propositions 2.3.2II and 2.3.3a imply that any purely temporal LSI transformation of a microbalanced random stimulus is microbalanced). However, $|\partial J/\partial t|$ (Fig. 2c) has most of its energy at those spatiotemporal frequencies whose velocity is equal to the velocity of the traveling contrast-reversal whose motion we wish to detect. Thus we see that, although J's motion cannot be exposed to standard analysis by a simple pointwise transformation, a temporal linear filter followed by a pointwise nonlinearity does suffice.

We turn now to the problem of stipulating the general conditions that a random stimulus I must satisfy so that $f \bullet I$ will be microbalanced for any pointwise transformation $f\bullet$. Call any random stimulus I *microbalanced under* a given transformation T iff $T(I)$ is microbalanced.

We state the following basic proposition (3.2) and its subsequent corollary (3.3) for continuously distributed random stimuli. The corresponding result for discretely distributed random stimuli is simpler and should be evident.

3.2. Necessary and Sufficient Conditions for a Random Stimulus to be Microbalanced under All Pointwise Transformations. *Let I be a random stimulus such that for any (x, y, t), $(x', y', t') \in \mathbf{Z}^3$, $(I[x, y, t], I[x', y', t'])$ has a continuous joint density. Then the following conditions are equivalent:*

(1) *I is microbalanced under all pointwise transformations.*

(2) *For all $x, y, t, x', y', t' \in \mathbf{Z}$, the joint density f of $(I[x, y, t], I[x', y', t'])$ and the joint density g of $(I[x, y, t'], I[x', y', t])$ satisfy*

$$f(p, q) + f(q, p) = g(p, q) + g(q, p) \tag{19}$$

for any $p, q \in \mathbf{R}$ such that $p \neq 0$ and $q \neq 0$.

Proof. Set $\kappa = I[x, y, t]$, $\lambda = I[x', y', t']$, $\gamma = I[x, y, t']$, and $v = I[x', y', t]$. Thus, (κ, λ) is distributed in $\mathbf{R}^2$ with density f and (γ, v) is distributed with density g.

((2) implies (1)): By definition of any pointwise transformation $h\bullet$, we have $h(0) = 0$. Thus we need integrate only over values of κ and λ which are both nonzero in computing the expectation $E[h(\kappa)\, h(\lambda)]$. In particular, if Eq. (19) is satisfied for all $p \neq 0$ and $q \neq 0$, then $h \bullet I$ is microbalanced since

$$E[h(\kappa)\,h(\lambda)] = \frac{1}{2}\left[\int_{\mathbf{R}}\int_{\mathbf{R}} h(p)\,h(q)\,f(p,q)\,dp\,dq + \int_{\mathbf{R}}\int_{\mathbf{R}} h(q)\,h(p)\,f(q,p)\,dq\,dp\right]$$
$$= \frac{1}{2}\left[\int_{\mathbf{R}}\int_{\mathbf{R}} h(p)\,h(q)\,f(p,q)\,dp\,dq + \int_{\mathbf{R}}\int_{\mathbf{R}} h(p)\,h(q)\,f(q,p)\,dp\,dq\right]$$
$$= \frac{1}{2}\int_{\mathbf{R}}\int_{\mathbf{R}} h(p)\,h(q)(f(p,q)+f(q,p))\,dp\,dq$$
$$= \frac{1}{2}\int_{\mathbf{R}}\int_{\mathbf{R}} h(p)\,h(q)(g(p,q)+g(q,p))\,dp\,dq = E[h(\gamma)\,h(\nu)]. \quad (20)$$

(Note: the boundedness and finite integrability of $h\bullet$ ensure that these expectations exist.)

(Not (2) implies not (1)): On the other hand, suppose Eq. (19) fails for some $x, y, t, x', y', t' \in \mathbf{Z}$. One way in which this might happen is if $f(r,r) > g(r,r)$ for some nonzero $r \in \mathbf{R}$. In this case, there exists a neighborhood N of r, not including 0, such that $f(m,n) > g(m,n)$ for all $m, n \in N$. Thus, for the function $h\colon \mathbf{R} \to \mathbf{R}$ defined by

$$h(n) = \begin{cases} 1 & \text{if } n \in N, \\ 0 & \text{otherwise,} \end{cases} \quad (21)$$

$h\bullet$ is a pointwise transformation (the function h is bounded on $\mathbf{R}$, finitely integrable, and $h(0) = 0$). However, $h\bullet I$ is not microbalanced since

$$E[h(\kappa)\,h(\lambda)] = \int_N\int_N f(m,n)\,dm\,dn > \int_N\int_N g(m,n)\,dm\,dn$$
$$= E[h(\gamma)\,h(\nu)]. \quad (22)$$

To recapitulate, if Condition 2 fails because there exists a nonzero $r \in \mathbf{R}$ for which $f(r,r) \neq g(r,r)$, then Condition 1 fails (I is not microbalanced under all pointwise transformations).

The only other way in which Condition 2 can fail is if $f(r,r) = g(r,r)$ for all $r \neq 0$ in $\mathbf{R}$, but for some $p, q \in \mathbf{R}$, with neither p nor q equal to 0, $f(p,q) + f(q,p) > g(p,q) + g(q,p)$. In this case, we obtain disjoint neighborhoods M of p and N of q, neither including 0, such that

$$f(m,n) + f(n,m) > g(m,n) + g(n,m) \quad (23)$$

for all $m \in M$, $n \in N$; consequently,

$$\int_M\int_N f(m,n) + f(n,m)\,dm\,dn > \int_M\int_N g(m,n) + g(n,m)\,dm\,dn. \quad (24)$$

Moreover, since—by assumption—$f(p,p) = g(p,p)$ and $f(q,q) = g(q,q)$, we can tailor the neighborhoods M and N to make the difference

$$\left[\int_M\int_M f(m,m')\,dm\,dm' + \int_N\int_N f(n,n')\,dn\,dn'\right]$$
$$-\left[\int_M\int_M g(m,m')\,dm\,dm' + \int_N\int_N g(n,n')\,dn\,dn'\right] \quad (25)$$

as small as we want. Consider, then, the function $h\colon \mathbf{R} \to \mathbf{R}$ defined by

$$h(u) = \begin{cases} 1 & \text{if } u \in M \cup N, \\ 0 & \text{otherwise.} \end{cases} \quad (26)$$

Again, $h\bullet$ is a pointwise transformation. However, $h\bullet I$ fails again to be microbalanced because, for suitably tailored M and N,

$$E[h(\kappa)\,h(\lambda)] = \int_M\int_M f(u,v)\,du\,dv + \int_N\int_N f(u,v)\,du\,dv$$
$$+ \int_M\int_N f(u,v) + f(v,u)\,du\,dv$$
$$> \int_M\int_M g(u,v)\,du\,dv + \int_N\int_N g(u,v)\,du\,dv$$
$$+ \int_M\int_N g(u,v) + g(v,u)\,du\,dv = E[h(\gamma)\,h(\nu)]. \quad ∎ \quad (27)$$

3.3. Corollary. *Let I be a random stimulus such that for all $(x, y, t), (x', y', t') \in \mathbf{Z}^3$, the pair $(I[x,y,t], I[x',y',t'])$ has a continuous joint density. Then I is microbalanced under all pointwise transformations if the following condition holds for all $x, y, t, x', y', t' \in \mathbf{Z}$: For f the joint density of $(I[x,y,t], I[x',y',t'])$, and g the joint density of $(I[x,y,t'], I[x',y',t])$, either*

$$f(p,q) = g(p,q) \qquad \textit{for all } p, q \in \mathbf{R},\ p \neq 0,\ q \neq 0, \quad (28)$$

or

$$f(p,q) = g(q,p) \qquad \textit{for all } p, q \in \mathbf{R},\ p \neq 0,\ q \neq 0. \quad (29)$$

Proof. If Eq. (28) holds for some $(x, y, t), (x', y', t') \in \mathbf{Z}^3$, then we also have

$$f(q,p) = g(q,p) \qquad \text{for all } p, q \in \mathbf{R},\ p \neq 0,\ q \neq 0, \quad (30)$$

and we obtain Eq. (19) by adding Eq. (28) and Eq. (30). The same reasoning applies for Eq. (29). ∎

A random stimulus microbalanced under all pointwise transformations, but quite different from J of example 3.1 is the following, suggested by J. Lappin (1989).

3.4. *Stimulus K: Rotating Random-Dot Cylinder.* Construct K by taking the parallel projection of a set of points on (and/or inside) the surface of a cylinder rotating around a vertical axis. Let the contrast values of the points be independent, identically distributed random variables. As is well known, when properly constructed, K can display a very strong kinetic depth effect, with dots moving in one direction seen as being in the front of the axis of rotation, and dots moving in the other direction seen as being in the back (Dosher, Landy, & Sperling, 1989; Ullman, 1979). Nonetheless, K is microbalanced under all pointwise transformations: All of K's systematic motion is horizontal; thus, we can drop reference to y, and note that for any x, t, x', t', the joint distribution of $(K[x, t], K[x', t'])$ is identical to that of $(K[x, t'], K[x', t])$. Hence, by Corollary 3.3, Condition 3, K is microbalanced under all pointwise transformations.

4. Texture Quilts

The rest of this paper is devoted to illustrating how the results of Section 3 can be applied to construct stimuli which display consistent apparent motion that cannot be exposed to standard analysis by any purely temporal transformation. Specifically, we demonstrate several motion-displaying stimuli, called *texture quilts* (Definition 4.1), that are microbalanced under all purely temporal transformations.

As illustrated in Fig. 3, the simplest transformations that suffice to expose the motion of texture quilts to standard analysis involve a purely spatial linear filter $s*$ followed by a rectifier $r\circ$:

$$T(Q) = r\circ(s * Q). \tag{31}$$

The spatial filter $s*$ will respond with varying energy throughout regions of the visual field, depending on whether or not the textures to which it is tuned populate those regions. However, the output of a linear filter to a texture is positive or negative depending on the local phase of the texture. The purpose of rectification is to transform regions of high-variance $s*$ response into regions of high average value, thus ensuring that the rectified output registers the presence or absence of texture, independent of phase. The result $T(Q)$ is a spatiotemporal function whose value reflects the local texture preferences of $s*$ in the visual field as a function of time (Bergen & Adelson, 1988; Caelli, 1985).[2]

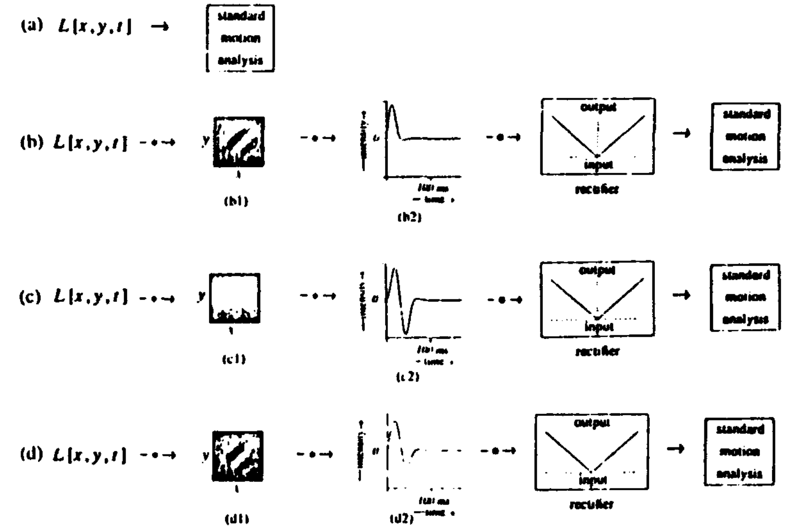


FIG. 3. Fourier and non-Fourier motion mechanisms. (a) Fourier motion mechanisms apply standard motion-analysis directly to the luminance signal L. (b), (c), and (d) Non-Fourier mechanisms apply standard motion-analysis to a nonlinear transformation of luminance. (b) A simple non-Fourier mechanism applies a signal transformation comprised of a spatiotemporal linear filter, followed by a *pointwise nonlinearity*. The $*$'s indicate spatial and temporal convolution, respectively, and $\circ$ indicates function composition. The filtering performed in (b) is roughly pointwise in time (the temporal impulse response b2 approximates an impulse), and the nonlinearity applied is a full-wave rectifier. This system (with appropriately chosen spatial filter, b1) will extract the motion of the texture quilts shown in Figs. 4b, 5d, 6c, and 6d. It will not extract the motion of stimulus J, the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a. (c) A spatially pointwise (the spatial impulse response c1 approximates an impulse), system with a flicker-sensitive temporal filter and a full-wave rectifier. Because of the flicker sensitivity, this mechanism will extract the motion of the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a but not the motion of the texture quilts shown in Figs. 4b, 5d, 6c, and 6d. (d) The temporal filter d2 averages the temporal filters b2 and c2, and the pointwise nonlinearity is a full-wave rectifier. With an appropriate spatial filter d1, the non-Fourier system extracts the motion of any corresponding texture quilt as well as the motion of the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a. However, it would be less well suited to these tasks than the detectors shown in (b) and (c) whose temporal filters it averages.

The essential trick in all the quilt examples we consider is to patch together various brief displays of static, random texture, taking appropriate measures to ensure that the resultant stimulus satisfies the following definition.

4.1. Definition of a Texture Quilt. Let $A \subset \mathbf{Z}^2$ be a set of points in space, and let $t_0, t_1, \ldots, t_N$ be a strictly increasing sequence of times, with $T = \{t \| t_0 \leqslant t < t_N\}$. Call any random stimulus Q satisfying the following conditions a *texture quilt*:

[2] In general, a spatial linear filter followed by a pointwise nonlinearity can have arbitrarily high order Volterra kernels, depending on the order of the Taylor series of the pointwise transformation. However, if we take the rectifier of step (2) to be $\mathrm{Rect}(x) = x^2$, then this squared output of a spatial filter is a second order spatial transformation. Standard motion analysis is yet another second order transformation. Thus, when we subject the squared filter output to standard motion analysis, we are applying a fourth order operator.

(i) Q assigns 0 to all points outside $A \times T$.

(ii) For $i = 0, 1, \ldots, N-1$, the random values assigned by Q to points in A at time t_i remain unchanged until time t_{i+1}.

(iii) *Independence*. For $i = 0, 1, \ldots, N-1$, the random substimuli Q^i, defined, for all points α in space and all times t, by

$$Q^i[\alpha, t] = \begin{cases} Q[\alpha, t] & t_i \leqslant t < t_{i+1}, \alpha \in A \\ 0 & \text{otherwise}, \end{cases} \tag{32}$$

are jointly independent.

(iv) *Symmetry*. For any $\alpha, \beta \in A$, and any $t \in T$, the joint distribution of $(Q[\alpha, t], Q[\beta, t])$ is identical to the joint distribution of $(Q[\beta, t], Q[\alpha, t])$.

Terminology. Call A and T respectively Q's *spatial* and *temporal regions of activity*, and for $i = 0, 1, \ldots, N-1$, call $\{t \| t_i \leqslant t < t_{i+1}\}$ the *i*th *timeblock* of Q.

The empirical usefulness of texture quilts derives from proposition 4.3 in conjunction with the fact that it is easy to construct various sorts of texture quilts which display consistent apparent motion across independent realizations. The proof of proposition 4.3 is eased by the following

4.2. Lemma. *Let Q be a texture quilt with spatial region of activity A. Then for any $\alpha, \beta \in A$, the pair of temporal functions (Q_α, Q_β) is distributed identically to the reverse pair (Q_β, Q_α).*

Proof. From Definition 4.1(i) and (ii), note that for temporal functions P and R, the density of the joint assignment $(Q_\alpha, Q_\beta) = (P, R)$ is 0 unless each of P and R is constant throughout each time block, and 0 outside T. Thus, any P and R for which the joint assignment $(Q_\alpha, Q_\beta) = (P, R)$ has nonzero density are completely determined by the values $P[t_i] = p_i$, and $R[t_i] = r_i$, for $i = 0, 1, \ldots, N-1$; for f_i the joint density of $(Q_\alpha[t_i], Q_\beta[t_i])$, Definition 4.1(iii) thus implies that the density of the joint assignment $(Q_\alpha, Q_\beta) = (P, R)$ is

$$\prod_{i=0}^{N-1} f_i(p_i, r_i). \tag{33}$$

But by Definition 4.1(iv), the quantity (33) is equal to

$$\prod_{i=0}^{N-1} f_i(r_i, p_i), \tag{34}$$

which is the density of the reverse occurrence that $(Q_\beta, Q_\alpha) = (P, R)$. ▌

4.3. Texture Quilts are Microbalanced under Purely Temporal Transformations. I. *Any texture quilt with a continuous joint density is microbalanced under all purely temporal, continuous transformations.*

II. *Any discretely distributed texture quilt is microbalanced under all purely temporal transformations.*

Proof of I. Let Q be a texture quilt with a continuous joint density, and let Φ be an arbitrary purely temporal, continuous transformation. We must prove that $\Phi(Q)$ is microbalanced. We can, of course, accomplish this by proving that $\Phi(Q)$ is microbalanced under all pointwise transformations (since, in particular, the identity transformation is pointwise). This turns out to be a convenient approach.

Let α, β be points in space, and let t and u be points in time. Because Φ is bounded and continuous and Q has a continuous joint density, we know that the joint density f of $(\Phi(Q)[\alpha, t], \Phi(Q)[\beta, u])$ and the joint density g of $(\Phi(Q)[\beta, t], \Phi(Q)[\alpha, u])$ both exist and are continuous on $\mathbf{R}^2$. We shall show for any $(p, r) \in \mathbf{R}^2$ with neither p nor r equal to 0, that either $f(p, r) = g(p, r)$ or $f(p, r) = g(r, p)$. The proposition will then follow from Corollary 3.3.

Case 1. At least one of α or β is outside A. Suppose α is outside A. Then by Definition 4.1(i), $Q_\alpha = 0$; hence $\Phi(Q)[\alpha, t] = \Phi(Q)[\alpha, u] = 0$. Consequently, $f(p, r) = g(r, p) = 0$ whenever $p \neq 0$. Thus Eq. (29) holds vacuously, with

$$f(p, r) = g(r, p) = 0 \qquad \text{for all } p, r \in \mathbf{R}, p \neq 0, r \neq 0. \tag{35}$$

Case 2. Both α and β are in A. Let F be the joint density of (Q_α, Q_β) and G be the joint density of (Q_β, Q_α). By Lemma 4.2, $F = G$. Clearly, then, for F_Φ the joint density of $(\Phi(Q_\alpha), \Phi(Q_\beta))$ and G_Φ the joint density of $(\Phi(Q_\beta), \Phi(Q_\alpha))$, it follows that $F_\Phi = G_\Phi$. For any $p, r \in \mathbf{R}$, recall that $f(p, r)$ is the density of the co-occurrence that $\Phi(Q)[\alpha, t] = p$, and $\Phi(Q)[\beta, u] = r$, but this is precisely the density of the event that $(\Phi(Q_\alpha)[t], \Phi(Q_\beta)[u]) = (p, r)$. This density, however, is equal to the integral of F_Φ over all pairs of temporal functions (P, R) such that $P[t] = p$ and $R[u] = r$. Similarly, $g(p, r)$ is the density of the co-occurrence that $\Phi(Q)[\beta, t] = p$, and $\Phi(Q)[\alpha, u] = r$, but this is the density of the event that $(\Phi(Q_\beta)[t], \Phi(Q_\alpha)[u]) = (p, r)$, which is equal to the integral of G_Φ over all pairs of temporal functions (P, R) such that $P[t] = p$ and $R[u] = r$. However, as we have already noted, $F_\Phi = G_\Phi$, implying that $f = g$. Apply Corollary 3.3 to complete the proof. ▌

The proof of II is similar.

The rest of Section 4 is devoted to showing how to construct two kinds of simple texture quilts. In Section 5, we apply these construction techniques in an experiment to investigate what sorts of textural characteristics are actually processed for motion information by the visual system.

4.4. *Binary Texture Quilts*

4.4.1. *A General Technique for Constructing Binary Texture Quilts.* The simplest sorts of texture quilts involve only two contrast values. As in Definition 4.1, let $T = \{t \| t_0 \leqslant t < t_N\}$ be the temporal region of activity, with new timeblocks beginning at times $t_0, t_1, \ldots, t_{N-1}$. Let A be the spatial region of activity. Associate with timeblocks $i=0, 1, \ldots, N-1$ spatial functions f_i (called *timeblock pictures*), each of which is 0 everywhere outside A, and takes only the values 1 and -1 within A. In addition, associate with timeblocks 0 through $N-1$ a family

$$\phi_0, \phi_1, \ldots, \phi_{N-1} \tag{36}$$

of jointly independent random variables, each of which takes the value 1 or -1 with equal probability. Then, for $i=0, 1, \ldots, N-1$, set

$$B_i[x, y, t] = \begin{cases} f_i[x, y] & \text{if } t \text{ is in timeblock } i, \\ 0 & \text{otherwise,} \end{cases} \tag{37}$$

and construct the random stimulus

$$B = \phi_0 B_0 + \phi_1 B_1 + \cdots + \phi_{N-1} B_{N-1}. \tag{38}$$

It is easy to see that B is a texture quilt. First, the functions B_i are defined to satisfy Definition 4.1(i) and (ii). The joint independence of the random variables ϕ_i ensures that B satisfies Definition 4.1(iii). To see that Definition 4.1(iv) is satisfied, note that for any $\alpha, \beta \in A$, either (i) $B_i[\alpha, t_i] = B_i[\beta, t_i]$ or (ii) $B_i[\alpha, t_i] = -B_i[\beta, t_i]$. In case (i),

$$B[\alpha, t_i] = \phi_i B_i[\alpha, t_i] = \phi_i B_i[\beta, t_i] = B[\beta, t_i], \tag{39}$$

implying that the pair $(B[\alpha, t_i], B[\beta, t_i])$ is distributed identically to the pair $(B[\beta, t_i], B[\alpha, t_i])$ (each pair with an equal probability of taking the value $(1, 1)$ or $(-1, -1)$). In case (ii)

$$B[\alpha, t_i] = -B[\beta, t_i], \tag{40}$$

and the pair $(B[\alpha, t_i], B[\beta, t_i])$ is distributed identically to the pair $(B[\beta, t_i], B[\alpha, t_i])$, each with an equal probability of assuming the value $(1, -1)$ or $(-1, 1)$. Thus Definition 4.1(iv) is satisfied along with 4.1(i), (ii), and (iii).

4.4.2. *Stimulus: The Sidestepping, Randomly Contrast-Reversing, Vertical Edge.* In Fig. 4b are displayed the 9 timeblock pictures comprising a particularly simple binary texture quilt. Note that the vertical dimension of Fig. 4b combines time and vertical space, precisely as a strip of movie film, scanned vertically, combines time and space. Timeblock pictures are separated by gray lines. Figure 4a shows the timeblock pictures f_0 through f_8 used in the construction. f_0 assigns the value -1 to all points (x, y) of the horizontal rectangle comprising the spatial region of activity, A. f_1 assigns 1 to the points in the leftmost eighth of A, and -1 to the points in the right seven-eighths. The timeblock pictures f_2 through f_8 continue to shift the vertical edge rightward through A until, in picture 8, A is uniformly 1. Multiplying each timeblock picture $i=1, 2, \ldots, 9$ by its associated random variable ϕ_i yields, in this particular realization, the stimulus given in Fig. 4b.

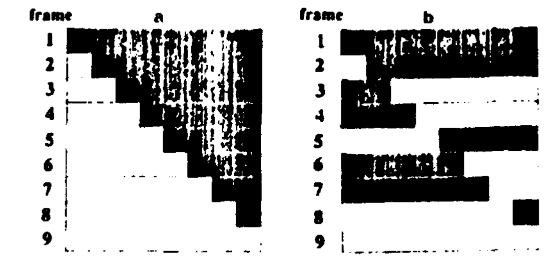


FIG. 4. Edge-driven motion from an ordinary edge and from a binary texture quilt. (a) A rightward moving light–dark edge visible to Fourier and non-Fourier motion systems. Nine entire frames are shown; each frame consists of an area of contrast +1 and area of contrast −1. (b) A realization of the *sidestepping, randomly contrast-reversing vertical edge*. This random stimulus is a texture quilt and hence microbalanced under all purely temporal transformations: that is, its rightward motion would be inaccessible to standard motion-analysis even if this analysis were preceded by an arbitrary, purely temporal transformation. Each frame of (b) was derived from the corresponding frame of (a) by multiplying the entire frame by a random variable that takes the value 1 or −1 with equal probability. The frame random variables are jointly independent. A straightforward way to extract the motion of this texture quilt is to (i) apply a linear filter sensitive to vertical edges, (ii) rectify the filtered output, and (iii) submit the result to standard motion analysis.

The construction of the sidestepping contrast-reversing edge (Fig. 4b) is symmetric to the construction of the traveling contrast-reversal of a random black-or-white vertical bar pattern (J in Fig. 2a). Transposing the x and t dimensions in Fig. 4b gives the xt-cross-section of a random stimulus J (e.g., Fig. 2a). This stimulus exhibits an unusual symmetry between space and time. Whereas the texture quilt of Fig. 4b is microbalanced under all purely temporal transformations, its transpose J (Fig. 2b) is microbalanced under all *purely spatial* transformations. Extracting motion from J requires *temporal* filtering followed by a nonlinearity. This process is essentially different from the process by which motion is extracted from texture quilts (e.g., Figs. 4b, 7a, 7b, and 7c) which requires a *spatial* nonlinearity.

4.4.3. *Stimulus: Oppositely Oriented Static Squarewaves Selected by a Drifting Grating.* Figure 5d shows the four timeblock pictures comprising another binary texture quilt constructed using technique 4.4.1. In Fig. 5a is shown a probabilistically defined sinewave grating, a stimulus whose motion is readily extracted by standard motion-analysis. In Figs. 5b1 and 5b2 are shown static vertical and horizontal squarewave gratings. The stimulus of Fig. 5c is obtained by using Fig. 5a to select between the vertical and horizontal gratings of Figs. 5b1 and 5b2. If the function of Fig. 5a is 1 at a certain point in space-time, the corresponding point in Fig. 5c is assigned the value of the corresponding point in Fig. 5b1; otherwise the point in Fig. 5c is assigned the value of the corresponding point in Fig. 5b2. Although Figs. 5c and 5d look similar, they differ in an important respect: the

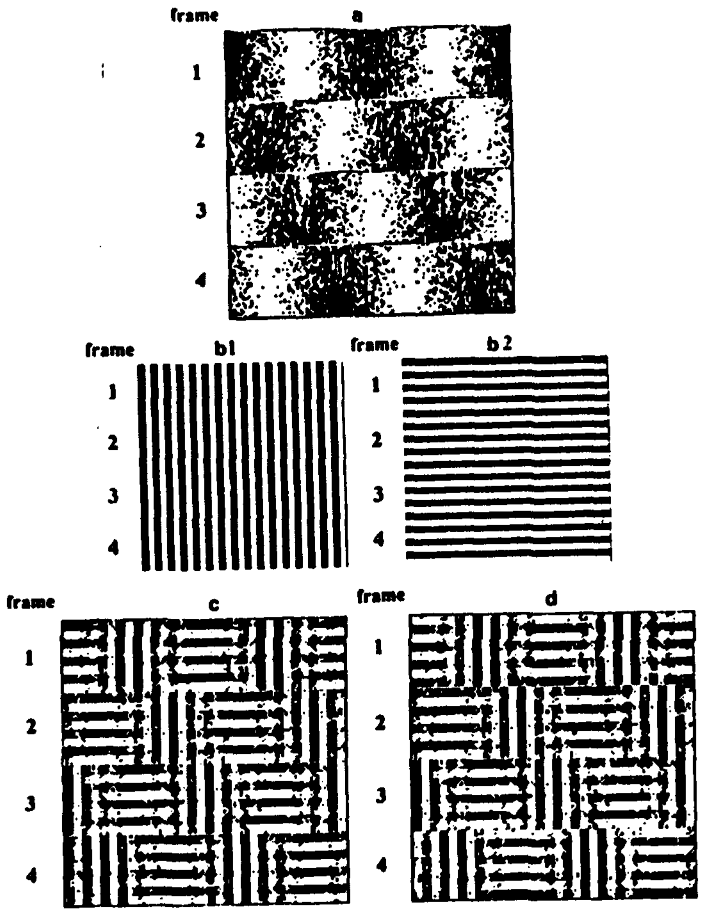


FIG. 5. Orientation-driven non-Fourier motion from a binary texture quilt. (a) A probabilistically defined sinewave grating that steps rightward 90 degrees between frames. The rightward motion in (a) *is accessible to all motion detectors.* (b1) Four frames of a static, vertical squarewave grating; (b2) Four frames of a static horizontal squarewave grating. (c) A rightward translating texture pattern. For every white point in (a), the corresponding value in (c) is chosen from the vertical squarewave grating in (b1); for every black point in (a), the corresponding value in (c) is chosen from the horizontal squarewave grating in (b2). (c) is not microbalanced; standard motion-analyzers can be designed to detect its motion. (d) A texture quilt. The frames of (d) are derived by multiplying the corresponding frames of (c) by jointly independent random variables, each of which takes the value 1 or −1 with equal probability. The texture quilt (d) is microbalanced under all purely temporal transformations, and therefore its rightward motion is unavailable *to any mechanism that applies standard motion analysis* to a purely temporal transformation of the visual signal.

stimulus of Fig. 5d is microbalanced under all purely temporal transformations, while that of Fig. 5c is not microbalanced. It is possible to design Fourier mechanisms to detect the motion of Fig. 5c, but not that of Fig. 5d. The critical difference is that the timeblock pictures of Fig. 5d are jointly independent, while those of Fig. 5c are not: Fig. 5d is obtained by randomly reversing the contrasts of the timeblock pictures of Fig. 5c.

4.5. *Sinusoidal Texture Quilts*

It is not difficult to elaborate technique 4.4.1 to a method for constructing quilts involving textures of arbitrarily many contrast values. We illustrate the principle in *the construction of quilts comprised of patches of sinusoidal grating.*

4.5.1. *A General Technique for Constructing Sinusoidal Texture Quilts.* As in Definition 4.1, let $T = \{t \| t_0 \leqslant t < t_N\}$ be the temporal region of activity, with new timeblocks beginning at times $t_0, t_1, ..., t_{N-1}$. Let A be the spatial region of activity. Associate with timeblocks $i = 0, 1, ..., N-1$, spatial functions W_i, each of which is C everywhere outside A, and takes only the values 1 and -1 within A. The stimulus in each time block will be composed of two components characterized by spatial frequencies (ω_i, θ_i) and $(\tilde{\omega}_i, \bar{\theta}_i)$, respectively, and independent phases $\rho_i, \tilde{\rho}_i$, respectively. Let

$$\omega_0, \theta_0, \tilde{\omega}_0, \bar{\theta}_0, \omega_1, \theta_1, \tilde{\omega}_1, \bar{\theta}_1, ..., \omega_{N-1}, \theta_{N-1}, \tilde{\omega}_{N-1}, \bar{\theta}_{N-1} \qquad (41)$$

be integers. Let P be an integer, and let

$$\rho_0, \tilde{\rho}_0, \rho_1, \tilde{\rho}_1, ..., \rho_{N-1}, \tilde{\rho}_{N-1} \qquad (42)$$

be jointly independent random variables, each uniformly distributed on the set $\{0, 1, ..., P-1\}$. Then, define the stimulus S as the sum of N component stimuli S_i defined in each timeblock,

$$S = \sum_{i=0}^{N-1} S_i, \qquad (43)$$

where, for $i = 0, 1, ..., N-1$, S_i is zero everywhere outside timeblock i; and for all t in timeblock i,

$$S_i[x, y, t] = f_i[x, y] = \begin{cases} \cos(2\pi(\omega_i x + \theta_i y - \rho_i)/P) & \text{if} \quad W_i[x, y] = 1, \\ \cos(2\pi(\tilde{\omega}_i x + \bar{\theta}_i y - \tilde{\rho}_i)/P) & \text{if} \quad W_i[x, y] = -1, \\ 0 & \text{otherwise.} \end{cases} \qquad (44)$$

It is easy to check that S satisfies Definition 4.1(i) and (ii). The joint independence of the random phase variables $\rho_i, \tilde{\rho}_i$, for $i = 0, 1, ..., N-1$ entails Definition 4.1(iii).

It remains to check that S satisfies Definition 4.1(iv). Consider points $\alpha, \beta \in A$. If $W_i[\alpha] \neq W_i[\beta]$, then, as is easily checked, $S[\alpha, t_i]$ and $S[\beta, t_i]$ are independent and identically distributed (each assuming a value from among $\{\cos(2\pi p/P) \| p = 0, 1, ..., P-1\}$ with equal probability). On the other hand, if $W_i[\alpha] = W_i[\beta]$, then the pair $(S[\alpha, t_i], S[\beta, t_i])$ is distributed identically to the pair $(S[\beta, t_i], S[\alpha, t_i])$ as a consequence of the following

Lemma. *Let $P \in \mathbf{Z}$, and let $\alpha = (\alpha_x, \alpha_y)$, $\beta = (\beta_x, \beta_y)$ and $\omega = (\omega_x, \omega_y)$ all be elements of $\mathbf{Z}^2$. Then for any integer $p \in \{0, 1, ..., P-1\}$, there exists an integer $q \in \{0, 1, ..., P-1\}$ such that (writing $\cdot$ for dot product)*

$$\cos(2\pi(\omega \cdot \alpha - p)/P) = \cos(2\pi(\omega \cdot \beta - q)/P) \tag{45}$$

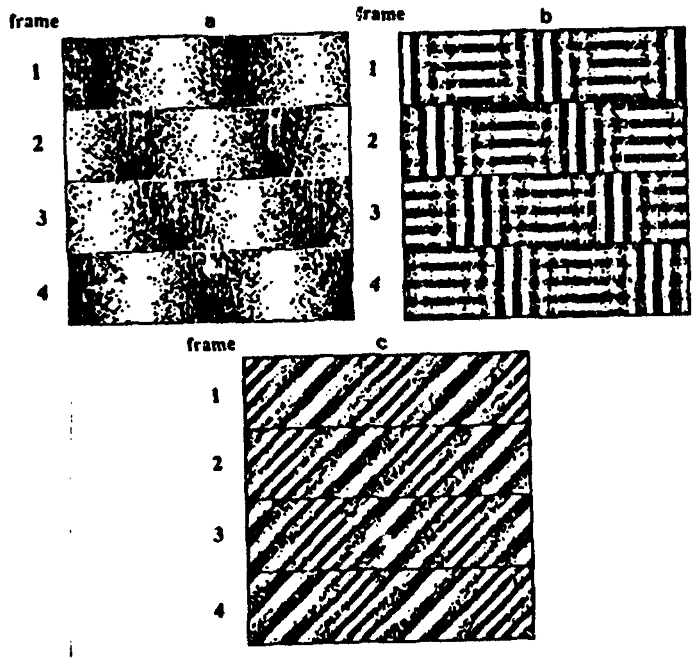


Fig. 6. Sinusoidal texture quilts: Motion driven by differences in *orientation* and in *spatial frequency*. (b) and (c) show realizations of random stimuli, each of which is microbalanced under all purely tem- poral transformations. Their rightward motion cannot be detected by any mechanism that applies standard motion analysis to a purely temporal transformation of the signal. In each case, the four frames in (a) select between two sinusoidal patterns. The phases of sinusoids are jointly independent across frames and across different-frequency sinusoidal components patched together in the same frame. The sinusoids mixed in (b) differ in orientation, whereas the sinusoids mixed in (c) have the same orientation, but differ in spatial frequency.

and

$$\cos(2\pi(\omega \cdot \beta - p)/P) = \cos(2\pi(\omega \cdot \alpha - q)/P). \tag{46}$$

Proof. As the reader may check, this is true for $q = (\omega \cdot \alpha + \omega \cdot \beta - p)$ modulo P. ∎

Thus, for α, β such that $W_i[\alpha] = W_i[\beta]$, we observe that for any outcome $\rho_i = p$, there exists an equally likely outcome $\rho_i = q$, such that

$$\begin{aligned}(\cos(2\pi(\omega_i \cdot \alpha - p)/P), \cos(2\pi(\omega_i \cdot \beta - p)/P)\\ = (\cos(2\pi(\omega_i \cdot \beta - q)/P), \cos(2\pi(\omega_i \cdot \alpha - q)/P)).\end{aligned} \tag{47}$$

We infer that the pair $(S[\alpha, t_i], S[\beta, t_i])$ is distributed identically to the pair $(S[\beta, t_i], S[\alpha, t_i])$.

4.5.2. *Stimulus: Oppositely Oriented Static Sinusoids Selected by a Drifting Grating.* The sinusoidal analog to the binary texture quilt of Fig. 5d is shown in Fig. 6b. In Fig. 6a are shown the functions W_1, W_2, W_3, and W_4 used to select between horizontal and vertical gratings. For this quilt, $\tilde{\omega}_i = \theta_i = 0$, for $i = 1, 2, 3, 4$; and for some integer F (with F/P the number of cycles per pixel), $\omega_i = \bar{\theta}_i = F$. The texture quilt of Fig. 6b modulates textural orientation across space and time. Alternatively, we can just as easily keep orientation constant and vary spatial frequency.

4.5.3. *Stimulus: Static Sinusoids of Different Spatial Frequencies, Selected by a Drifting Grating.* Figure 6c shows a texture quilt using the sampling functions of Fig. 6a, but setting $\omega_i = \theta_i = 2\tilde{\omega}_i = 2\bar{\theta}_i$ for $i = 1, 2, ..., 4$.

5. What Aspects of Texture Does the Visual System Process for Motion?

In this section, we describe a psychophysical experiment investigating the question of what characteristics of spatial texture are analyzed for motion information by the visual system. Three texture quilts are compared across four different viewing conditions. These conditions comprise a sequence of similar but increasingly challenging motion discrimination tasks.

5.1. *Procedure*

Every texture quilt used in this experiment is comprised of a sequence of jointly independent timeblocks, each lasting 1/30 s. (Each timeblock consists of two identical refreshes at 1/60 s.) Each texture quilt is stochastically periodic with a period of 8 timeblocks: that is, for any integer i, the ith timeblock is identically distributed to the $i+8$th timeblock. Accordingly, we refer to 8 timeblocks of the texture quilt as one *cycle*. The motion elicited by each quilt is carried by a squarewave that *selects between two textures, and steps 1/4 cycle on every odd timeblock*. The squarewave thus completes one of its four-step cycles in each 8 timeblock cycle of the quilt.

On each trial, a texture quilt moving randomly left or right is presented, and the subject is required to signal (with a button-press) which way the quilt appeared to move. The subject is asked to maintain fixation on a small spot present in the middle of the stimulus throughout the display, and receives feedback after each trial. For each quilt under each viewing condition, the subject performs 100 practice trials followed directly by 100 actual trials. Quilt realizations are jointly independent across trials. The starting phase of the quilt is chosen randomly on each trial.

The Four Viewing Conditions. For a given quilt, the four viewing conditions differ with respect to the number of quilt cycles displayed. In Condition 1, the easiest condition, the subject sees two quilt cycles (each cycle comprised of eight stimulus timeblocks), with each timeblock displayed for 1/30 s. In Conditions 2, 3, and 4, the subject sees 1.5, 1, and 0.5 quilt cycles, respectively.

5.1.1. *Three Quilt Stimuli.* The first quilt (the F-quilt) modulates textural spatial frequency as a function of space and time, while keeping orientation constant. The 8 timeblocks comprising one full cycle of the F-quilt are shown in Fig. 7a. A second quilt (the O-quilt, Fig. 7b) modulates textural orientation as a function of space and time, while keeping spatial frequency constant. A third quilt (the E-quilt, Fig. 7c) spatiotemporally modulates texture between jointly independent binary noise and the so-called "even" texture (Julesz, Gilbert, & Victor, 1978).

All stimuli were viewed from 1 m against a mean luminant background. At this distance, each quilt spanned 6.8 horizontal and 3.2 vertical degrees, and the modulating squarewave moved at an average velocity of 12.75 deg/s.

5.1.2. *Why These Three Quilts.* In each of the three quilts, a squarewave with vertical bars is used to modulate between two textures as a function of space and time. The squarewave has a spatial frequency of 0.3 c/deg, and steps 1/4 cycle rightward on every odd timeblock (temporal frequency 3.75 Hz, velocity 12.75 deg/s). We use a 1/4 cycle stepping squarewave to modulate between the two textures comprising each quilt in order to rule out the possibility that the motion elicited by the quilt is being carried by the border between textural regions. That is, the 1/4 cycle stepping squarewave has the advantage that the signal derived from the borders between texture regions is ambiguous in motion content. Given the requirement of 1/4 cycle steps, we changed the particular instantiation of the quilt on even timeblocks (i.e., within steps of the squarewave) in order to spread textural energy broadly in temporal frequency without altering the spatial frequency content of the texture.

It has been previously observed (Green, 1986; Ramachandran, Ginsburg, & Anstis, 1983; Watson & Ahumada, 1983a) that motion is carried more effectively by spatiotemporal variation of textural spatial frequency than by variation of textural orientation. The F-quilt and O-quilt were chosen to further investigate this claim. The E-quilt is of interest because the two textures of which it is composed (jointly independent binary noise and the even texture) have identical second order

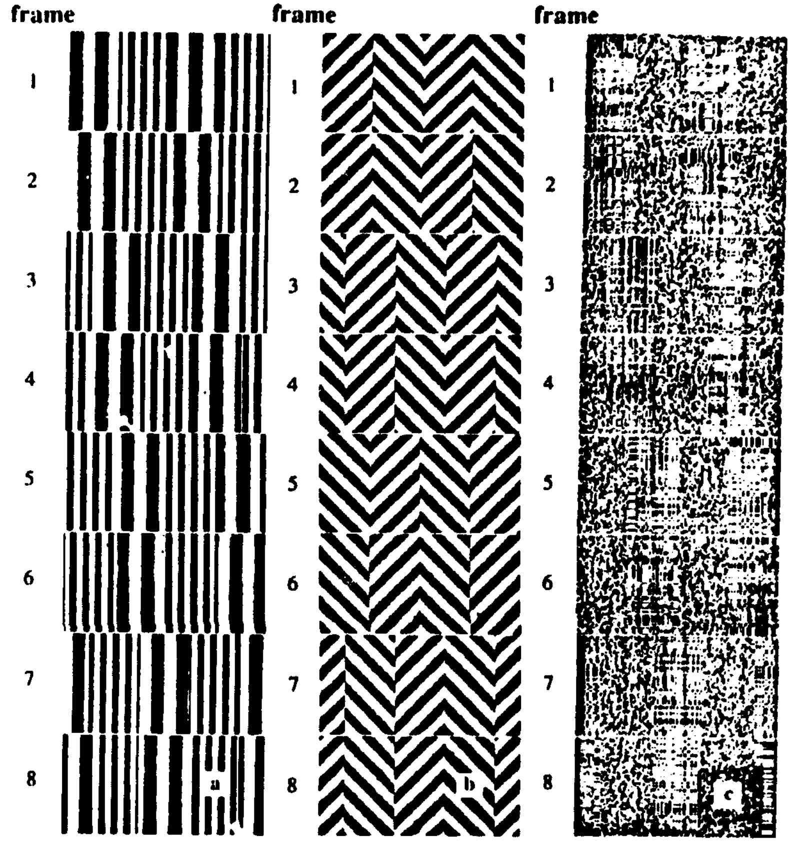


FIG. 7. Three quilts used to study motion carried by modulation of texture spatial frequency, by texture orientation, and by higher order textural characteristics. (a) Eight frames that comprise one cycle of the F-quilt. Motion is generated by a squarewave modulation of textural spatial frequency. The squarewave grating selects between vertical sinusoidal gratings of spatial frequency 1.2 and 2.4 c/deg. The texture-modulating squarewave is 0.3 c/deg, and steps 1/4 cycle rightward on every odd frame. Every even frame is independent of and distributed identically to the preceding frame. Presentation proceeds at the rate of 30 frames/s. This gives the texture-modulating squarewave a temporal frequency of 3.75 Hz and a mean velocity of 25 deg/s. (b) Eight frames that comprise one cycle of the O-quilt. In the O-quilt, textural orientation is modulated by the same squarewave used to modulate spatial frequency in the F-quilt. The O-quilt squarewave selects between oppositely oriented sinusoidal gratings that have a spatial frequency of 2.8 c/deg. (c) Eight frames that comprise one cycle of the E-quilt. In the E-quilt, the texture-modulating squarewave selects between jointly independent binary noise and an even texture (Julesz, Gilbert, & Victor, 1978). Despite the evident difference between these two textures, every time-independent linear filter has the same expected power for both textures. Thus, if motion-from-texture resulted from applying a simple squaring transformation to the output of a spatial linear filter and submitting the result to standard motion analysis, the motion of the E-quilt would be invisible.

statistics. That is, the joint distribution of any given pair of points in space is the same under both the component textures of the E-quilt. This means that, despite the obvious difference in appearance between the component textures, the expected energy in the response of any given spatial linear filter is the same for both component textures. If the pointwise nonlinearity applied to the output of the spatial linear filter prior to motion analysis were simple squaring, it would be impossible to detect the motion of the E-quilt.

Victor and Conte (1990) studied apparent motion elicited by E-quilts, and noted that it is much weaker than motion elicited by comparable stimuli (also texture quilts) that modulate between textures differing in spatial frequency. Our experiment confirms this finding.

5.2. *Results*

Two subjects participated in the study, CC (the experimenter) and GA (naive). The results for CC are shown in Fig. 8 bottom, and those for GA are shown in Fig. 8 top. Note first that both subjects were able to reliably discriminate left/right motion in all three stimuli although subject GA failed with the E-quilt at the briefest exposure. The two subjects performed comparably well at motion direction discrimination of the O-quilt, but CC was much better than GA at detecting the motion of both the F-quilt and the E-quilt. Subject CC was better at detecting the motion of the F-quilt than the O-quilt; the reverse was true of subject GA.

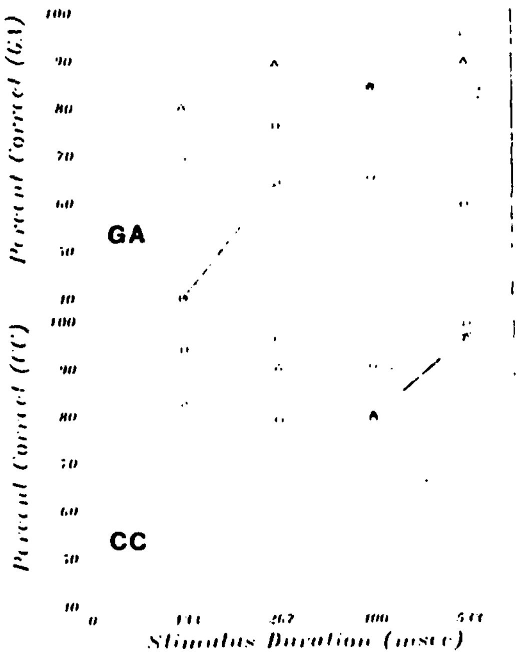


FIG. 8. The percent of correct direction-of-motion judgments to the F-quilt, the O-quilt, and the E-quilt as a function of stimulus duration. The panels show data for subjects CC and GA, respectively. Each data point is the mean of 100 judgments. (Squares) F-quilt; (triangles) O-quilt; (circles) E-quilt. The stimulus durations of 133, 266, 400, and 533 ms, correspond to stimulus presentations of 0.5, 1, 1.5, and 2 quilt cycles.

It is possible that these performance differences reflect a genuine differences in the perceptual apparatus of the two subjects. However, we cannot *rule out the* possibility that the better performance of subject CC is due merely to his vastly greater experience with motion perception tasks of this sort.

5.3. *Discussion*

Many of the models proposed to explain rapid, preattentive segregation of spatial textures (Beck, Sutter, & Ivry, 1987; Bergen & Adelson, 1988; Caelli, 1985; Malik & Perona, 1989; Sutter, Beck, & Graham, 1989) can easily be adapted to deal with the motion displayed by texture quilts. The texture segregation models in this class typically subject the visual input function to a linear transformation (a "texture grabber") followed by a pointwise nonlinearity (such as a rectifier or thresholder) to indicate the presence or absence of the texture. Such models propose that two contiguous textural regions would generate a perceptual boundary if the visual system were equipped with a linear filter that is differentially tuned to one of the textures.

An analogous mechanism to detect the motion of texture quilts, suggested by the current experiment and the work of Victor and Conte (1990), (*i*) convolves the input stimulus with a spatial texture-grabbing filter tuned to the moving texture, then (ii) squares the output of the filter, to transform regions of high energy filter output into regions of high average value, and (iii) subjects the rectified output to standard motion analysis. However, the transformation applied in steps (i) and (ii) does not distinguish between the two textures comprising the E-quilt, and therefore fails to account for the good performance with the E-quilt. A simple modification to deal with texture segregation and motion perception of the E-quilt is to assume some other post-filter rectification operation than the squaring operation. It is quite easy to choose a linear filter in combination with a post-filter rectifier (other than the squaring operation) that will segregate the random and even textures (e.g., Julesz & Bergen, 1983). The current experiment does not specifically indicate the kind of rectification that might be involved.

What sorts of filters are available to the visual system to compute motion from texture? For example, Daugman (1985) points out that (i) Gabor filters provide an optimal tradeoff between resolution in the space and spatial frequency domains, and (ii) many investigators note that simple cells in cat striate cortex are well modeled by oriented Gabor filters (e.g., Andrews & Pollen, 1979; DeValois, DeValois, & Yund, 1979; Wilson & Sherman, 1976). Are the linear filters that serve motion-from-texture computations Gabor-like cortical simple cells? The theory reported here provides a tool, and the demonstration experiments illustrate how it might be used to answer such questions.

6. Summary

The main contributions of this paper are to (i) introduce the notion of a random stimulus *microbalanced under all pointwise transformations*, (ii) provide necessary and sufficient conditions for a random stimulus to be of this sort, (iii) use this result to construct apparent motion stimuli called *texture quilts* that are microbalanced under all purely temporal transformations, and (iv) show that subjects can reliably discriminate the motion direction of three kinds of texture quilts.

Texture quilts provide a flexible array of tools for studying motion perception that is truly mediated by spatiotemporal modulation of spatial texture without contamination by mechanisms responsive to the motion extracted directly by standard analysis or motion extracted by standard analysis of any purely temporal transformation of the stimulus.

Acknowledgments

The research reported here was supported by USAF Life Science Directorate, Visual Information Processing Program *Grants 85-0364 and 88-0140*.

References

Adelson, E. H., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2), 284–299.

Adelson, E. H., & Bergen, J. (1986). The extraction of spatio-temporal energy in human and machine vision. *Proceedings of the IEEE Workshop on Motion: Representation and Analysis*, 151–155.

Andrews, B. W., & Pollen, D. A. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology (London)*, 287, 163–176.

Anstis, S. M. (1970). Phi movement as a subtraction process. *Vision Research*, 10, 1411–1430.

Baker, C. L., & Braddick, O. (1982a). Does segregation of differently moving areas depend on relative or absolute displacement. *Vision Research*, 22, 851–856.

Baker, C. L., & Braddick, O. (1982b). The basis of area and dot number effects in random dot motion perception. *Vision Research*, 22, 1253–1260.

Beck, J., Sutter, A., & Ivry, R. (1987). Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing*, 37, 299–325.

Bell, H. H., & Lappin, J. S. (1979). The detection of rotation in random dot patterns. *Perception and Psychophysics*, 26, 415–417.

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature*, 333(6171), 363–364.

Bowne, S. F., McKee, S. P., & Glaser, D. A. (1989). Motion interference in speed discrimination. *Journal of the Optical Society of America A*, 6(7), 1112–1121.

Braddick, O. (1973). The masking of apparent motion in random-dot patterns. *Vision Research*, 13, 355–359.

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research*, 14, 519–527.

Caelli, T. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision*, 1(1), 19–30.

Cavanagh, P. (1988). Motion: The long and the short of it. Presented at *Conference on Visual Form and Motion Perception: Psychophysics, Computation, and Neural Networks* (Meeting dedicated to the memory of the late Kvetoslav Prazdny). Boston University, MA, March 5, 1988.

Cavanagh, P., Arguin, M., & von Grunau, M. (1989). Interattribute apparent motion. *Vision Research*, 29(9), 1197–1204.

Chang, J. J., & Julesz, B. (1983a). Displacement limits, directional anisotropy and direction versus form discrimination in random dot cinematograms. *Vision Research*, 23, 639–646.

Chang, J. J., & Julesz, B. (1983b). Displacement limits for spatial frequency random-dot cinematograms in apparent motion. *Vision Research*, 23, 1379–1386.

Chang, J. J., & Julesz, B. (1985). Cooperative and non-cooperative processes of apparent movement of random dot cinematograms. *Spatial Vision*, 1(1), 39–45.

Chubb, C., & Sperling, G. (1987). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Investigative Ophthalmology and Visual Science*, 28, 233.

Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, 5(11), 1986–2007.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7), 1160–1169.

Derrington, A. M., & Badcock, D. R. (1985). Separate detectors for simple and complex grating patterns? *Vision Research*, 25, 1869–1878.

Derrington, A. M., & Henning, G. B. (1987). Errors in direction-of-motion discrimination with complex stimuli. *Vision Research*, 27, 61–75.

DeValois, K. K., DeValois, R. L., & Yund, E. W. (1979). Responses of striate cortical cells to grating and checkerboard patterns. *Journal of Physiology (London)*, 291, 483–505.

van Doorn, A. J., & Koenderink, J. J. (1984). Spatiotemporal integration in the detection of coherent motion. *Vision Research*, 24, 47–54.

Dosher, Barbara A., Landy, M. S., & Sperling, G. (1989). Ratings of kinetic depth in multi-dot displays. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 116–4,5.

Green, M. (1986). What determines correspondence strength in apparent motion. *Vision Research*, 26, 599–607

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: Univ. of Chicago Press.

Julesz, B., & Bergen, J. R. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *Bell Systems Technical Journal*, 62(6), 1619–1645.

Julesz, B., Gilbert, E., & Victor, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*, 31, 137–140.

Lappin, J. S. (1989). Personal communication, June 20.

Lappin, J. S., & Bell, H. H. (1972). Perceptual differentiation of sequential visual patterns. *Perception and Psychophysics*, 12, 129–134.

Lelkins, A. M. M., & Koenderink, J. J. (1984). Illusory motion in visual displays. *Vision Research*, 24, 1083–1090.

Malik, J., & Perona, P. (1989). *A computational model of texture perception* (Computer Science Division (EECS) Report No. UCB/CSD 89/491). Berkeley: University of California.

Marr, D. & Ullman, S. (1981). Direction selectivity and its use in early visual processing. *Proc. R. Soc. London, Ser. B*, 211, 151–180.

Nakayama, K., & Silverman, G. (1984). Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, 24, 293–300.

Pantle, A., & Turano, K. (1986). Direct comparisons of apparent motions produced with luminance, contrast-modulated (CM), and texture gratings. *Investigative Ophthalmology and Visual Science*, 27(3), 141.

Petersik, J. T., Hicks, K. I., & Pantle, A. J. (1978). Apparent movement of successively generated subjective figures. *Perception*, 7, 371-383.

Ramachandran, V. S., & Anstis, S. M. (1983). Displacement thresholds for coherent apparent motion in random dot-patterns. *Vision Research*, 23, 1719–1724.

Ramachandran, V. S., Ginsburg, A., & Anstis, S. M. (1983). Low spatial frequencies dominate apparent motion. *Perception*, 12, 457–461.

Ramachandran, V. S., Rao, V. M., & Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research*, 13, 1399–1401.

van Santen, J. P. H., & Sperling, G. (1984). A temporal covariance model of motion perception. *Journal of the Optical Society of America A*, 1, 451–473.

van Santen, J. P. H., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, 2(2), 300–321.

Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3, 263–346, 1984.

Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods and Instrumentation*, 8, 144–151.

Sutter, A., Beck, J., & Graham, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception and Psychophysics*, 46(4), 312-332.

Turano, K., & Pantle, A. (1989). On the mechanism that encodes the movement of contrast variations: Velocity discrimination. *Vision Research*, 29(2), 207-221.

Ullman, S. (1979). *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.

Victor, J. D., & Conte, M. M. (1990). Motion mechanisms have only limited access to form information. *Vision Research*, 30(2), 289-301.

Watson, A. B., & Ahumada, A. J., Jr. (1983a). A linear motion sensor. *Perception*, 12, A17.

Watson, A. B., & Ahumada, A. J., Jr. (1983b). *A look at motion in the frequency domain*. NASA Technical Memorandum 84352.

Watson, A. B., & Ahumada, A. J., Jr. (1985). A model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2(2), 322-342.

Watson, A. B., Ahumada, A. J., & Farrell, J. E. (1986). The window of visibility: A psychophysical theory of fidelity in time-sampled motion displays. *Journal of the Optical Society of America A*, 3(3), 300-307.

Wilson, J., & Sherman, S. (1976). Receptive field characteristics of neurones in cat striate cortex: Changes with visual field eccentricity. *Journal of Neurophysiology*, 39, 512-533.

Received: February 8, 1989

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium

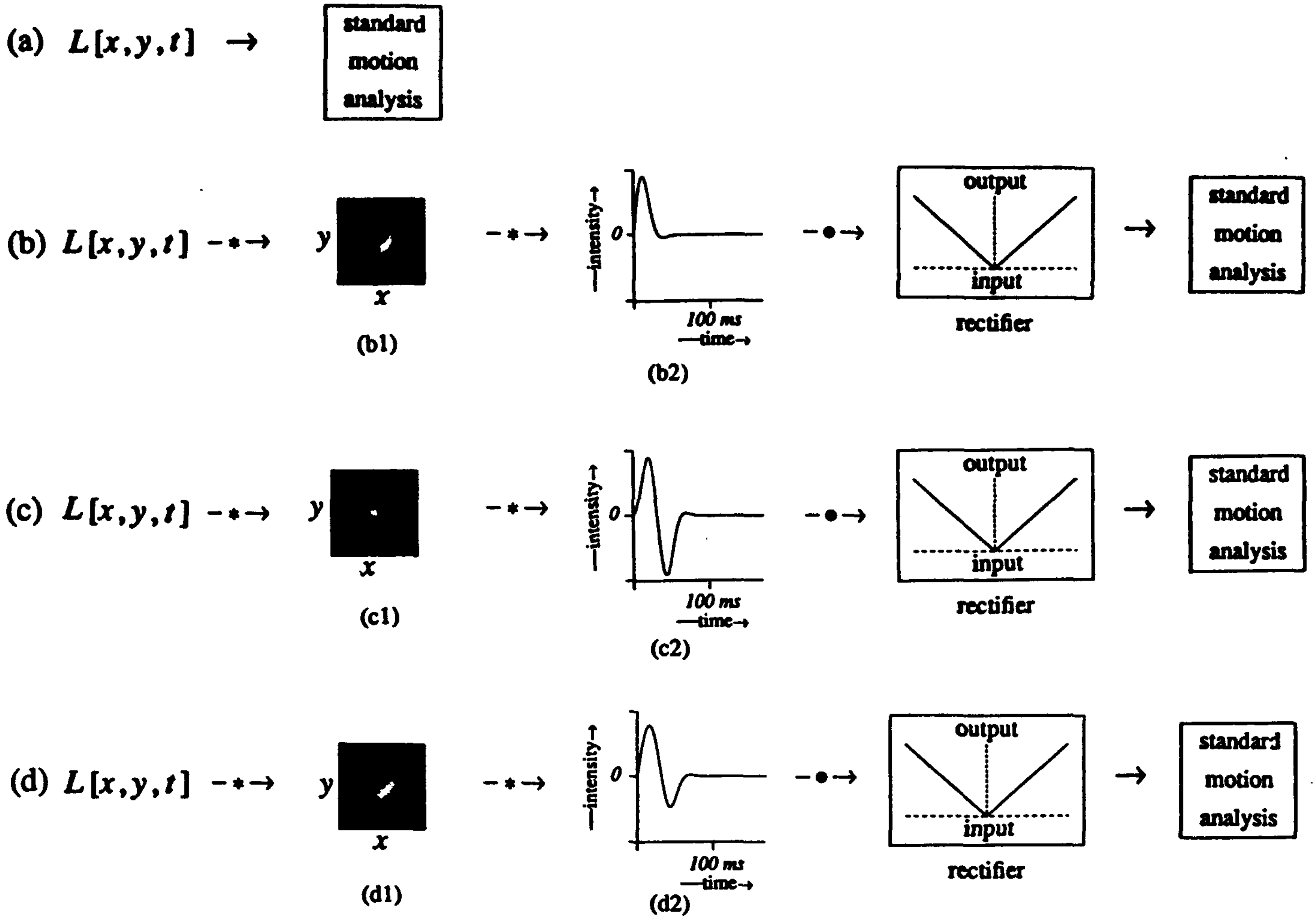


Fig. 3. Fourier and nonFourier motion mechanisms. (a) Fourier motion mechanisms apply standard motion-analysis directly to the luminance signal L. (b, c, d) NonFourier mechanisms apply standard motion analysis to a nonlinear transformation of luminance. (b) A simple nonFourier mechanism applies a signal transformation comprised of a spatiotemporal linear filter, followed by a pointwise nonlinearity. The ∗ 's indicate spatial and temporal convolution, respectively, and • indicates multiplication. The filtering performed in (b) is roughly pointwise in time (the temporal impulse response b2 approximates an impulse), and the nonlinearity applied is a full-wave rectifier. This system (with appropriately chosen spatial filter, b1) will extract the motion of the texture quilts shown in Figs. 4b, 5d, 6c, and 6d. It will not extract the motion of stimulus J, the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a. (c) A spatially pointwise (the spatial impulse response c1 approximates an impulse), system with a flicker-sensitive temporal filter and a full-wave rectifier. Because of the flicker sensitivity, this mechanism will extract the motion of the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a but not the motion of the texture quilts shown in Figs. 4b, 5d, 6c, and 6d. (d) The temporal filter d2 averages the temporal filters b2 and c2, and the pointwise nonlinearity is a full-wave rectifier. With an appropriate spatial filter d1, ths nonFourier system extracts the motion of any corresponding texture quilt as well as the motion of the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a. However, it would be less well-suited to these tasks than the detectors shown in (b) and (c) whose temporal filters it averages.

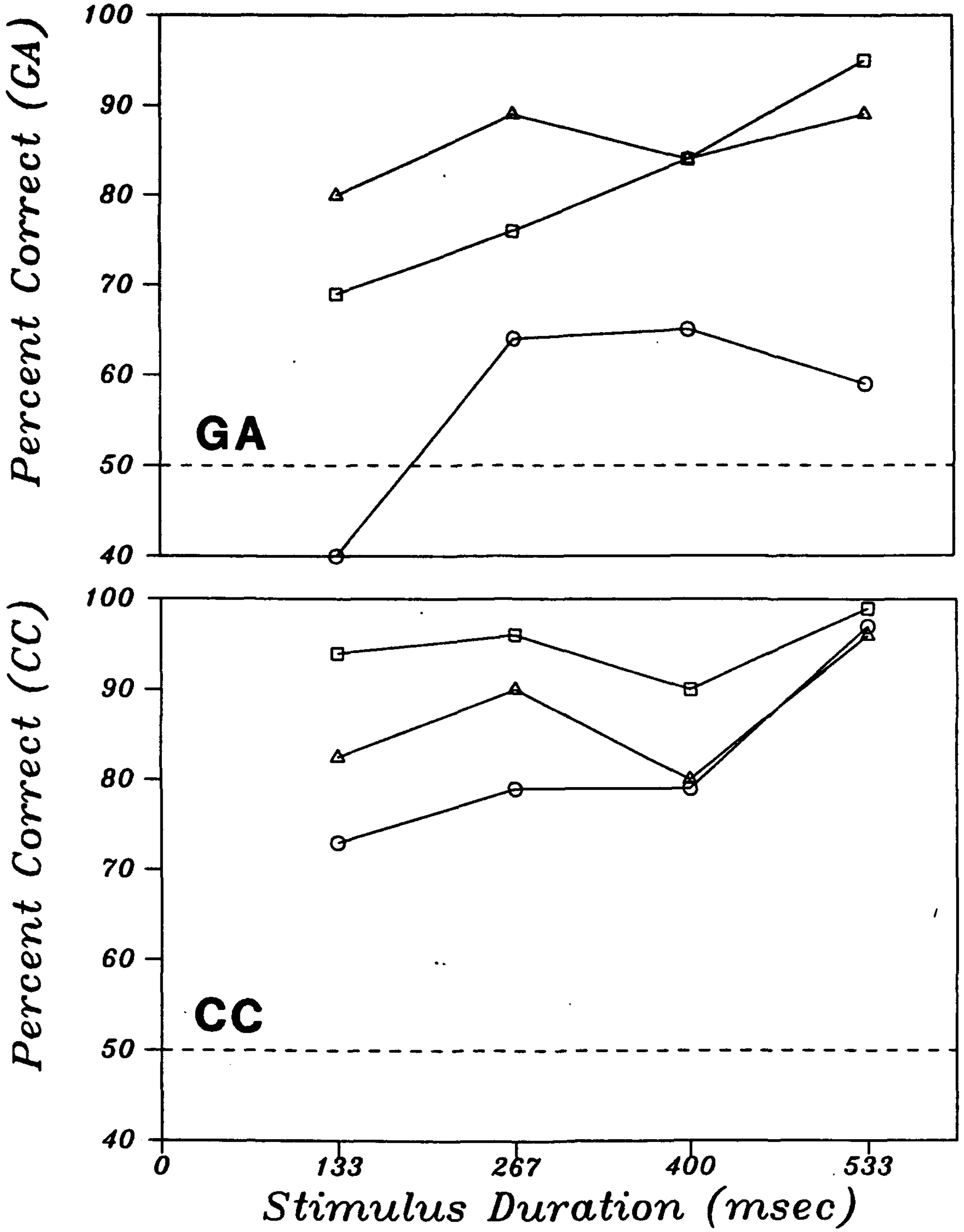


Fig. 8. The percent of correct direction-of-motion judgments to the F-quilt, the O-quilt, and the E-quilt as a function of stimulus duration. The panels show data for subjects CC and GA, respectively. Each data point is the mean of 100 judgments. (Squares) F-quilt; (triangles) O-quilt; (circles) E-quilt. The stimulus durations of 133, 266, 400, and 533 ms, correspond to stimulus presentations of 0.5, 1, 1.5 and 2 quilt cycles.

Vision Res. Vol. 31, No. 7/8, pp. 1399–1415, 1991
Printed in Great Britain. All rights reserved

0042-6989/91 $3.00 + 0.00

OBJECT SPATIAL FREQUENCIES, RETINAL SPATIAL FREQUENCIES, NOISE, AND THE EFFICIENCY OF LETTER DISCRIMINATION

DAVID H. PARISH and GEORGE SPERLING*

Human Information Processing Laboratory, Department of Psychology and Center for Neural Sciences, New York University, NY 10003, U.S.A.

(Received 7 July 1988; in revised form 2 June 1990)

Abstract—To determine which spatial frequencies are most effective for letter identification, and whether this is because letters are objectively more discriminable in these frequency bands or because we can utilize the information more efficiently, we studied the 26 upper-case letters of English. Six two-octave wide filters were used to produce spatially filtered letters with 2D-mean frequencies ranging from 0.4 to 20 cycles per letter height. Subjects attempted to identify filtered letters in the presence of identically filtered, added Gaussian noise. The percent of correct letter identifications vs *s/n* (the root-mean-square ratio of signal to noise power) was determined for each band at four viewing distances ranging over 32:1. Object spatial frequency band and *s/n* determine *presence of information* in the stimulus; viewing distance determines retinal spatial frequency, and affects only *ability to utilize*. Viewing distance had no effect upon letter discriminability: object spatial frequency, not retinal spatial frequency, determined discriminability. To determine discrimination efficiency, we compared human discrimination to an ideal discriminator. For our two-octave wide bands, *s/n* performance of humans and of the ideal detector improved with frequency mainly because linear bandwidth increased as a function of frequency. Relative to the ideal detector, human efficiency was 0 in the lowest frequency bands, reached a maximum of 0.42 at 1.5 cycles per object and dropped to about 0.104 in the highest band. Thus, our subjects best extract upper-case letter information from spatial frequencies of 1.5 cycles per object height, and they can extract it with equal efficiency over a 32:1 range of retinal frequencies, from 0.074 to more than 2.3 cycles per degree of visual angle.

Spatial filtering Scale invariance Psychophysics Contrast sensitivity Acuity

INTRODUCTION

Characterizing objects

When we view objects, what range of spatial frequencies is critical for recognition, and how is our visual system adapted to perceive these frequencies? Ginsburg (1978, 1980) was among the first to investigate this problem by means of spatial bandpass filtered images of faces and lowpass filtered images of letters. He noted the lowest frequency band for faces and the cutoff frequency for letters at which the images seemed to him to be clearly recognizable. The cutoff frequency for letters was 1–2 cycles per letter width; faces were best recognized in a band centered at 4 cycles per face width. He also proposed that the perception of geometric visual illusions, such as the Mueller–Lyer and Poggendorf, was mediated by low spatial frequencies (Ginsberg, 1971, 1978; Ginsberg & Evans, 1979).

An issue that is related to the lowest frequency band that suffices for recognition is the encoding economy of a band. For a filter with a bandwidth that is proportional to frequency (e.g. a two-octave-wide filter), the lower the frequency, the smaller the number of frequency components needed to encode the filtered image of a constant object. Combining these two notions, Ginsburg concluded that objects were best, or most efficiently, characterized by the lowest band of spatial frequencies that sufficed to discriminate them. Ginsburg (1980) went on to suggest that higher spatial frequencies were redundant for certain tasks, such as face or letter recognition.

Several investigators were quick to point out that objects can be well discriminated in various spatial frequency bands. Fiorentini, Maffei and Sandini (1983) observed that faces were well recognized in either high or in lowpass filtered bands. Norman and Erlich (1987) observed that high spatial frequencies were essential for discrimination between toy tanks in photographs.

*To whom reprint requests should be addressed.

With respect to geometric illusions, both Janez (1984) and Carlson, Moeller and Anderson (1984) observed that the geometric illusions could be perceived for images that had been highpass filtered so that they contained no low spatial frequencies. This suggests that low and high spatial frequency bands may carry equivalently useful information for higher visual processes.

Characterizing the visual system

In the studies cited above, the discussion of spatial filtering focuses on *object* spatial frequencies, that is, frequencies that are defined in terms of some dimension of the object they describe (cycles per object). Most psychophysical research with spatial frequency bands has focused on *retinal* spatial frequencies, that is, frequencies defined in terms of retinal coordinates. For example, the spatial contrast sensitivity function (Davidson, 1968; Campbell & Robson, 1968) describes the threshold sensitivity of the visual system to sine wave gratings as a function of their *retinal* spatial frequency. Visual system sensitivity is greatest at 3–10 cycles per degree of visual angle (c/deg). How does visual system sensitivity relate to object spatial frequencies?

Unconfounding retinal and object spatial frequencies

Retinal spatial frequency and object spatial frequency can be varied independently to determine whether certain object frequencies are best perceived at particular retinal frequencies. Object frequency is manipulated by varying the frequency band of bandpass filtered images; retinal frequency is manipulated by varying the viewing distance.

The cutoff *object* spatial frequency of lowpass filters and the observer's viewing distance were varied independently by Legge, Pelli, Rubin and Schleske (1985) who studied reading rate of filtered text at viewing distances over a 133:1 range. Over about a 6:1 middle range of distances, reading rate was perfectly constant, and it was approximately constant over a 30:1 range. At the longest viewing distances, there was a sharp performance decrease (as the letters became indiscriminably small). At the shortest viewing distance, performance decreased slightly, perhaps due to large eye movements that the subjects would have to execute to bring relevant material towards their lines of sight, and to the impossibility of peripherally previewing new text.

While viewing distance changed the overall level of performance in Legge et al., the cutoff *object* frequency of their low-pass filters at which performance asymptoted did not change. From this study, we learn that reading rate can be quite independent of retinal frequency over a fairly wide range, and that dependence on critical object frequency does not depend on viewing distance. Because the authors measured reading rate only in lowpass filtered images, we cannot infer reading performance in higher spatial frequency bands from their data.

Unconfounding object statistics and visual system properties

Human visual performance is the result of the combined effects of the objectively available information in the stimulus, and the ability of humans to utilize the information. In studying visual performance with differently filtered images, it it critical to separate availability from ability to utilize. For example, narrow-band images can be completely described in terms of a small number of parameters—Fourier coefficients or any other independent descriptors—than wide-band images. Poor human performance with narrow-band images may reflect the impoverished image rather than an intrinsically human characteristic—an ideal observer would exhibit a similar loss.

The problem of assessing the utility of stimulus information becomes acute in comparing human performance in high and in low frequency bandpass filtered images. Typically, filters are constructed to have a bandwidth proportional to frequency (constant bandwidth in terms of octaves). For example, Ginsburg (1980) used faces filtered into 2-octave-wide bands; while Norman and Ehrlich (1987) also used 2-octave bands for their filtered tank pictures. With such filters, high spatial frequency images contain more independent frequencies than low frequency images.

Although linear bandwidth represents perhaps the important difference between images filtered in octave bands at different frequencies, the informational content of the various bands also depends critically on the nature of the specific class of objects, such as faces or letter. Obviously, determining the information content of images is a difficult problem. When it is not solved, the amount of stimulus information available within a frequency band is confounded with the ability of human observers to use the information. Direct comparisons of performance between differently filtered objects are inappropriate. This distinction between objectively available stimulus information and the human ability to use it has not been adequately posed in the context of spatial bandpass filtering.

Efficiency

In the present context, physically available information is best characterized by the performance of an ideal observer. If there were no noise in the stimulus, the ideal observer would invariably respond perfectly. To compare the performance of an observer, human or ideal, noise of root-mean-square (r.m.s.) amplitude n is progressively added to the signal of r.m.s. amplitude s until the performance is reduced to some criterion, such as 50% correct in a letter identification task. This defines the signal to noise ratio, $(s/n)_c$, for a criterion c. Efficiency *eff* of human performance is defined by:

$$eff = \left(\frac{s_i}{n_i}\right)_c^2 \Big/ \left(\frac{s_h}{n_h}\right)_c^2$$

where h and i indicate *human* and *ideal* observers, and s and n are r.m.s. signal and noise amplitudes (Tanner & Birdsall, 1958). In a pure, quantally limited system, efficiency actually represents the fraction of quanta absorbed (utilization efficiency). In the context of signal detection theory, efficiency is given by a d' ratio:

$$eff = (d'_h/d'_i)^2.$$

Overview

For an object that contains a broad spectrum of spatial frequencies, object spatial frequency is determined by the center frequency of a spatial bandpass filtered image. Retinal spatial frequency is determined by the viewing distance at which the stimulus is viewed. Stimulus information is determined jointly by the signal-to-noise ratio, by the spatial filtering, and by the characteristics of the set of signals; these three informational components are combined in the efficiency computation. Letters are a convenient stimulus to study because they are highly overlearned so that human performance can be expected to be reasonably efficient, and because much is already known about the visibility of letters in the presence of internal noise (letter acuity) and about the visual processing of letters.

Specifically, to determine the roles of object and retinal spatial frequencies, letters are filtered into various frequency bands. Noise is added, and the psychometric function for correct identification is determined as a function of s/n. Accuracy depends only on s/n and not on overall contrast, for a wide range of contrasts (Pavel, Sperling, Riedl & Vanderbeck, 1987). This determination is repeated for every combination of object frequency band and viewing distance. Thereby, retinal spatial frequency and object spatial frequency are unconfounded, enabling us to determine whether a particular object frequency band is better discriminated in one visual channel (retinal frequency) than any other (Parish & Sperling, 1987a, b). Moreover, by computing an ideal observer for the identification task, we obtain an objective measure of the information that is present in each of the frequency bands. Finally, the comparison of human performance with the performance of the ideal observer gives us a precise measure of the ability of our subjects to utilize the information in the stimulus. Having untangled these factors, we can determine which spatial frequencies most efficiently characterize letters for identification.

METHOD

Two experiments were conducted using similar stimuli and procedures.

Stimuli

Letters (signals) and noise. The original, unfiltered letters were selected from a simple 5 × 7 upper-case font commonly used on CRT terminals. Since this is an experiment in pattern recognition, we felt that the simplest letter pattern might be the most general; indeed, this font has been widely used in letter discrimination studies. For the purpose of subsequent spatial filtering, the letters were redefined on a pixel grid that measured 45 (vertical height) × 35 (maximum horizontal extent of letters M and W). The letters had value 1 (white); the background had value 0 (black). To avoid edge effects in filtering, the background was extended to 128 × 128 pixels for all computations. However, only the center 90 × 90 pixels of the stimulus were displayed, as these contained effectively all the usable stimulus information, even for low spatial-frequency stimuli. Letters for presentation were chosen pseudo-randomly from the set of 26 upper-case English letters. Noise

Table 1. Parameters of the bandpass filters: lower and upper half-amplitude frequencies, peak, and 2D mean frequencies in cycles/letter height

Band	Lower	Peak	Upper	Mean*
0	0	Lowpass	0.53	0.39
1	0.26	0.53	1.05	0.74
2	0.53	1.05	2.11	1.49
3	1.05	2.11	4.22	2.92
4	2.11	4.22	8.44	5.77
5	6.33	Highpass	22.5	20.25

*Frequencies are weighted according to their squared amplitude (power) in computing the mean.

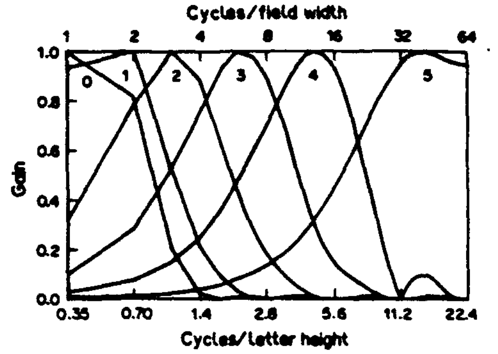


Fig. 1. Filter characteristics for the filters used in the experiments. There are two abscissas, both on a log scale. The top abscissa is the frequency in cycles per unwindowed field width (128 pixels); the bottom abscissa is in cycles per letter height (45 pixels). The ordinate is the normalized gain. The parameter i indicates the filter designation b_i in the text.

fields were defined on a 128 × 128 array by choosing independent Gaussian noise samples for each pixel, with the mean equal to zero and a variance σ^2 as required by the condition. (As with the letters, only the central 90 × 90 pixels were displayed.) Forty different noise fields were created.

Filters. Each stimulus consisted of a filtered letter added to an identically filtered noise field. Six spatial filters were available, corresponding to six successive levels of a Laplacian pyramid (Burt & Adelson, 1983). The zero-frequency component was added to the images so that they could be viewed. The object-relative filter characteristics, upper and lower half-amplitude cutoff and 2D mean frequency (cycles per letter height), appear in Table 1. The 2D mean frequency $\bar{f}$ for a given band is:

$$\bar{f} = \sum_{x=0}^{127} \sum_{y=0}^{127} f_{x,y} a_{x,y}^2 \Big/ \sum_{x=0}^{127} \sum_{y=0}^{127} a_{x,y}^2$$

where $f_{x,y}$ is the 2D frequency and $a_{x,y}$ is its amplitude. Cycles per object height is used rather than the more usual cycles per object width because the height of our upper-case letters remained constant across the entire set, whereas the width varied between letters.

The transfer functions (spectra) of the filters are displayed in Fig. 1. Approximately, filters are separated in spatial frequency by an octave (factor of 2) and have a bandwidth at half-amplitude of two octaves. The small mound in the lower right corner of Fig. 1 is a negligible imperfection in filter 4. For convenience, the limited range of spatial frequencies passed by each of the filters will be referred to as the *band* of that filter; a specific band is b_i ($i = 0, 1, 2, 3, 4, 5$), where b_0 is the lowest set of frequencies and b_5 is the highest.

The filter spectra (shown in Fig. 1) are approximately symmetrical in log frequency coordinates, a symmetrical spectrum in log coordinates is highly skewed to the right in linear frequency coordinates, resulting in a mean that is much greater than the mode. In a 2D (vs 1D) filter, the rightward shift is accentuated. For example, band 2 has a peak frequency of 1.05 c/object but a 2D mean frequency of 1.49 c/object. The single most informative characterization of such a skewed bandpass spectrum depends somewhat on the context; usually use the mean rather than the peak.

Figure 2 (top) shows the letter G, filtered in bands 1–5 without noise; the bottom shows the same signals plus noise, $s/n = 0.5$. The full 128 × 128 array (extended by reflection beyond its edges) was passed through the filter so that the effect of the picture boundary did not intrude into the critical part of the display.

Signal to noise ratio, s/n. A filtered letter is a *signal*. Let i, j index a particular pixel in the x, y coordinate space of the stimulus. The signal contrast $c_s(i,j)$ of pixel i, j is:

$$c_s(i,j) = \frac{(l(i,j) - l_0)}{l_0} \qquad (1)$$

where $l_{i,j}$ is the luminance of pixel i, j and l_0 is the mean signal luminance over the 90 × 90 array. Signal power per pixel, s, is defined as mean contrast power averaged over the 90 × 90 pixel array:

$$s = (IJ)^{-1} \sum_{i}^{I} \sum_{j}^{J} c_s(i,j)^2 \qquad (2)$$

where $c_{i,j}$ is the contrast of pixel i, j and $I = J = 90$.

Noise contrast $c_n(i,j)$ is the value of the i,jth noise sample divided by the mean luminance. Analogously to signal power (equation 2), noise contrast power per pixel, n, is equal to $(\sigma/l_0)^2$. The signal to noise ratio is simply s/n.

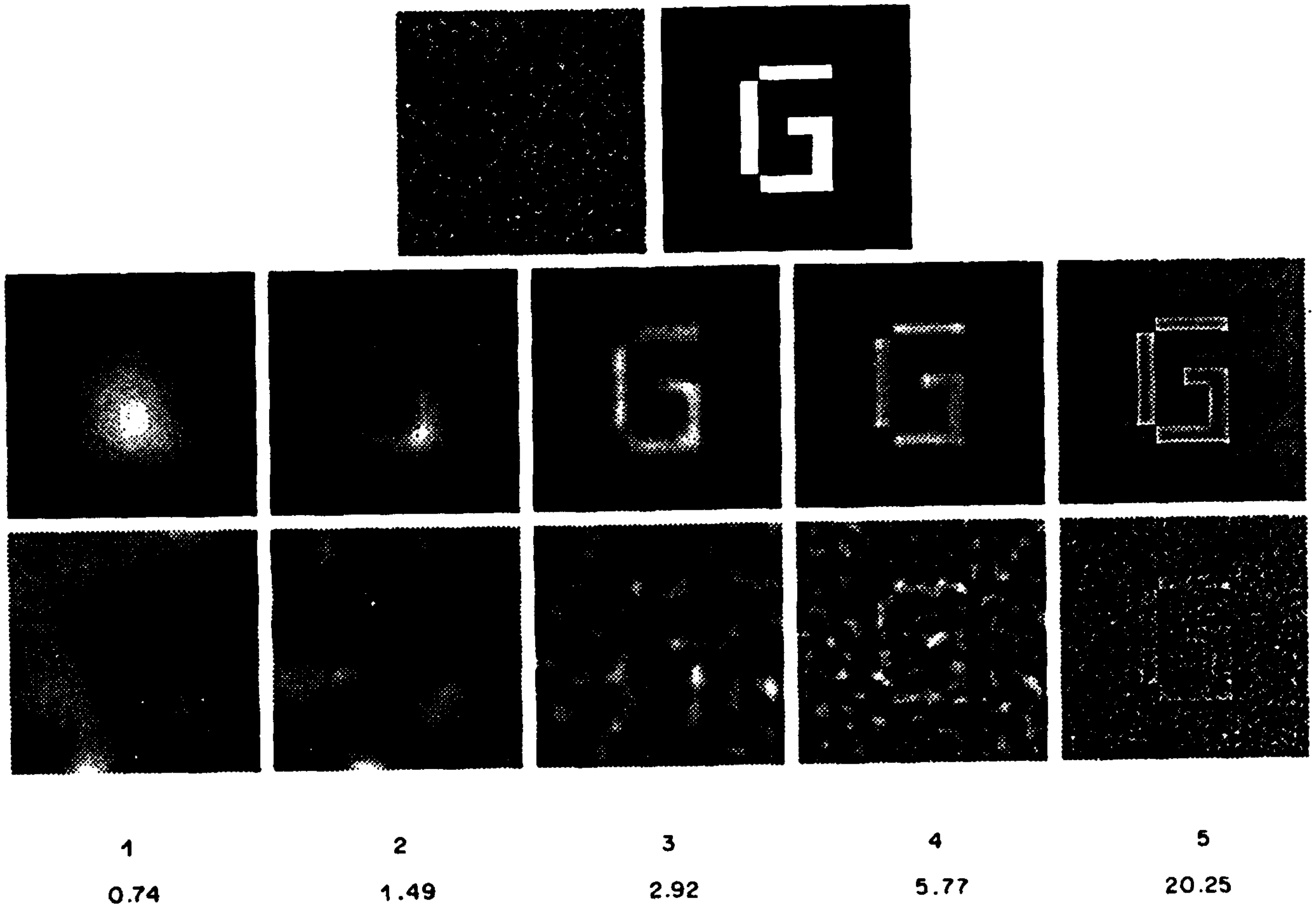


Fig. 2. Top: unfiltered noise and unfiltered letter G. Middle: the letter G filtered in spatial frequency bands 1 5 with only quantization noise. Bottom: filtered letter G plus filter noise in the same bands with a signal-to-noise ratio of 0.50 in all panels. The effective s/n in the reproduction is somewhat lower (from Parish & Sperling, 1987a). The first row of numerals indicates the number by which the filter band is referred to in the text; the bottom row indicates the *mean* frequency of the bands in cycles per letter height.

Quantization. Our display system produced 256 discrete luminance levels. Level 128 was used as the mean luminance l_0; l_0 was 47.5 cd/m^2. To produce a visual display of a given letter, band, and s/n, signal power s and noise power n were normalized so that the luminance of every one of the 8100 displayed pixels fell within the range of the display system; there was no truncation of the tails of the Gaussian noise. (Although the relationship between input gray-level and output luminance was not quite linear at the extreme intensity values, it was determined that more than 90% of the pixels fell within the linear intensity range.) Intensity normalization was applied separately to each stimulus (combination of signal plus noise). By normalizing the total stimulus $s + n$, the actual value of s displayed to the subject diminished as n increased; i.e. the actual value of s was not known by the subject. Indeed, even stimuli with precisely the same letter in the same band and with the same s/n might be produced with slightly different s and n depending on the extreme values of the noise fields.

Seven values of s/n were available for each band, chosen in a pilot study to insure that the data yielded the entire psychometric function (chance to best performance). The same pilot study showed that subjects never performed above chance when confronted with noise-free letters from b_0; this band was omitted from the present study.

Procedure: experiment 1

Four of the experimental variables—letter identity, noise field, frequency band, and s/n—were randomized within each session. A fifth variable, viewing distance, was held constant within each session and was varied between sessions. Four viewing distances were used: 0.121, 0.38, 1.21 and 3.84 m. A chin rest was used to stabilize the subject's head for viewing at the shortest distance. At the four distances, the 90 × 90 pixel stimulus subtended 31.6, 10, 3.16 and 1.0 deg of visual angle respectively. The upper and lower half-amplitude cut-off retinal frequencies for the upper six filters, with respect to the four viewing distances used in this experiment, and for a fifth distance used in the second experiment, appear in Table 2. Subjects participated in four 1-hr sessions at each viewing distance. Each session consisted of 315 trials, nine trials at each of seven s/n's for each of the five frequency bands.

Prior to the first session, subjects were shown noise-free examples of the unfiltered letters. They were told that each stimulus presentation consisted of a letter and a certain amount of noise, and that the letter may appear degraded in some way. They were informed that at no time would a letter be shifted in orientation or from its central location in the stimulus field. Finally, they were instructed to view each stimulus for as long as they desired before making their best guess as to which letter had been presented. A response (letter identity) was required on every trial. Subjects typed the response on a keyboard connected to the host computer (Vax 11/750); subsequently, typing a carriage return erased the video screen and initiated the next trial in a few seconds. The room illumination was very dim; the response keyboard was lighted by stray light from its associated CRT terminal. No feedback was offered to the subjects.

Observers

Three subjects, two male and one female, between the ages of 20 and 27 participated in the experiment. All subjects had normal or corrected-to-normal vision. One of the subjects was a paid participant in the study.

Procedure: experiment 2

This experiment was run before expt 1. It is reported here because it offers additional data with two new and one old subject at a fifth viewing distance. Except as noted, the procedures are similar to expt 1. The screen was viewed through a darkened hood at a distance

Table 2. Lower and upper half-power frequency and 2D mean frequency (in c/deg of visual angle) for all bands and viewing distances used in both experiments

	Viewing distance (m)				
Band	0.12	0.38	1.21	3.84	0.48
0 (lowpass)	0.00–0.04 (0.03)	0.00–0.12 (0.09)	0.00–0.37 (0.27)	0.00–1.18 (0.87)	0.00–0.15 (0.11)
1	0.02–0.07 (0.05)	0.06–0.23 (0.16)	0.18–0.74 (0.52)	0.58–2.34 (1.65)	0.07–0.29 (0.21)
2	0.04–0.15 (0.10)	0.12–0.47 (0.33)	0.37–1.48 (1.04)	1.18–4.70 (3.30)	0.15–0.59 (0.41)
3	0.07–0.30 (0.20)	0.23–0.94 (0.64)	0.74–2.97 (2.04)	2.34–9.40 (6.48)	0.29–1.18 (0.81)
4	0.15–0.59 (0.40)	0.47–1.88 (1.27)	1.48–5.94 (4.04)	4.70–18.80 (12.82)	0.59–2.36 (1.60)
5 (highpass)	0.30–2.25 (1.41)	0.94–7.13 (4.45)	2.97–22.53 (14.19)	9.40–71.27 (45.00)	1.77–8.96 (5.63)

of 0.48 m. At this distance, the 90 × 90 stimuli subtended 7.15 deg of visual angle. The half-amplitude cut-off frequencies and the mean frequencies of the six spatial filters are given in the rightmost column of Table 2. Three male subjects between the ages of 20 and 27 participated in the experiment. All subjects had normal or corrected-to-normal vision. Two of the subjects were paid for their participation, and one, DHP, also participated in expt 1. Five sessions of 315 trials were run for each subject.

RESULTS

Psychometric functions: $\hat{p}$ vs $\log_{10} s/n$

The measure of performance is the observed probability $\hat{p}$ of a correct letter identification. The complete psychometric functions are displayed in Figs 3 (expt 1) and 4 (expt 2). A separate psychometric function is shown for each subject, viewing distance and frequency band. In band b_1, for all subjects, performance asymptotes (for noiseless stimuli) at $\hat{p} \approx 0.5$. In all other bands, performance improves from near-chance (1/26) to near perfect as the value of s/n increases.

Noise resistance as a function of frequency band

An obvious aspect of the data of both experiments is that the data move to the left of the figure panels as band spatial frequency increases. This means that high spatial frequency stimuli (bands b_4, b_5) are identifiable at smaller

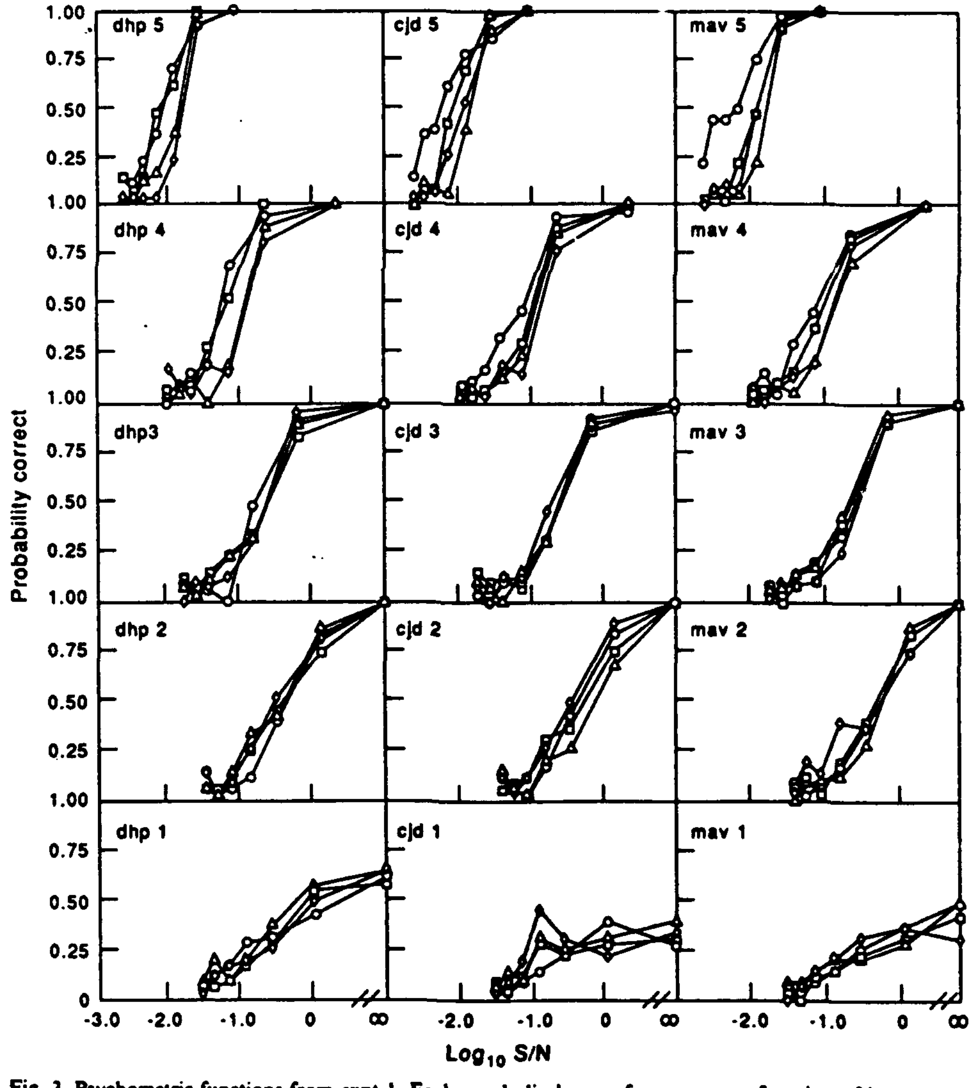


Fig. 3. Psychometric functions from expt 1. Each graph displays performance as a function of $\log_{10} s/n$, within a frequency band. The parameter is viewing distance. Subjects are arranged in columns and frequency band is arranged in rows, progressing from the highest frequency band at the top to the lowest band at the bottom. The four viewing distances are 3.84 (○), 1.21 (△), 0.38 (□), and 0.121 (◇) m.

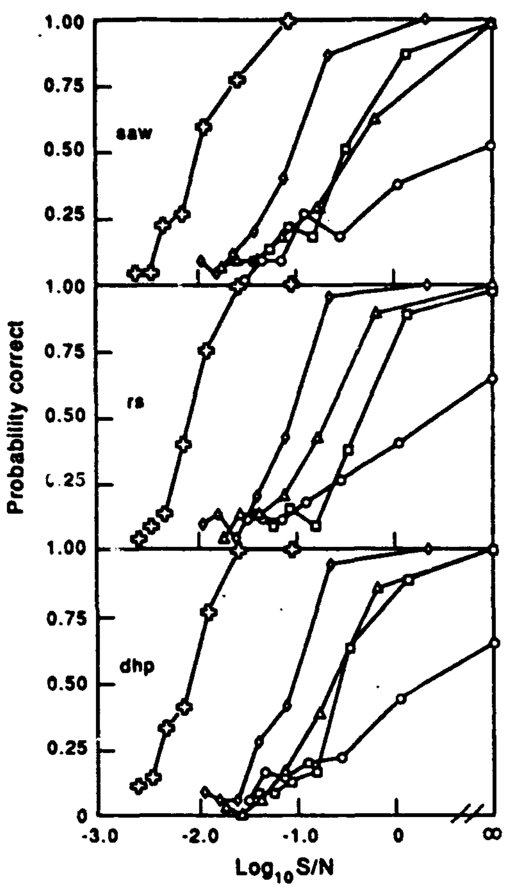


Fig. 4. Psychometric functions for each subject and frequency band in expt 2. Viewing distance was 0.48 m. The five frequency bands, b_1–b_5, are indicated, respectively, by ○, □, △, ◇ and +. The probability of a correct response is plotted as a function of $\log_{10} s/n$.

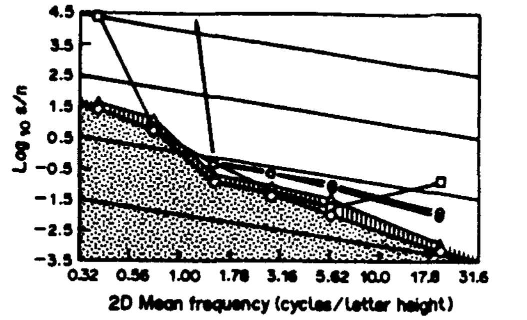


Fig. 5. Performance of human subjects and various computational discriminators. The abscissa indicates $\log_{10}$ of the mean frequency of each bandpass stimulus. The ordinate indicates the (interpolated) s/n ratio at which a probability of a correct response $p = 0.5$ is achieved. Circles indicate each of the three subjects in expt 1 at the intermediate viewing distance of 1.21 m. In band b_1, 2 of 3 human subjects fail to achieve 50% correct (*eff* = 0); these points lie outside the graph. (△) indicates sub-ideal and (◇) indicates super-ideal performances of discriminators that brackets the ideal discriminator. The shaded area below the super-ideal discriminator indicates theoretically unachievable performance. Squares indicate performance of a spatial correlator-discriminator. The oblique parallel lines have slope −1 that represents the improvement in expected performance (decrease in s/n) as function of the number of frequency components in each band when filter bandwidth is proportional to frequency.

s/n than stimuli in bands b_1 and b_2; resistance to noise increases with spatial frequency band. To enable comparisons of noise sensitivity as a function of band, the s/n at which $\hat{p} = 50\%$ was estimated for each subject and frequency band from expt 1 by means of inverse interpolation from the best fitting logistic function. As viewing distance had no effect, all estimates were made using the data collected when viewing distance was equal to 0.38 m. A graph of these $(s/n)_{50\%}$ points as a function of the mean object frequency of the band is plotted in Fig. 5 (○). For comparison, the expected rate of improvement in $(s/n)_{50\%}$, based on the increasing number of frequency components as one moves from low to high frequency bands, is plotted as a series of parallel lines in Fig. 5. Performance improves [$(s/n)_{50\%}$ decreases] somewhat faster than $1/f$ (the slope of the parallel lines). These results, and Fig. 5, will be analyzed in detail in the Discussion section.

The non-effect of viewing distance

Another property of the data is that, in most conditions, viewing distance has no effect on performance. Analysis of variance, carried out individually for each subject, shows that there is no significant effect of distance in any band for subject dhp and a significant effect of distance in bands b_4 and b_5 for the other two subjects. Further analysis by a Tukey test (Winer, 1971) in bands b_4 and b_5 for these subjects shows that the only significant effect of distance is that visibility at the longest viewing distance is *better* than at the other three distances. For subject CJD, the improvement is equivalent to a gain in s/n of 0.19 and 0.28 $\log_{10}$ (for bands b_4 and b_5, respectively); for MAV, the corresponding gains were 0.21 and 0.40.

Improved performance at long viewing distances is almost certainly due to the square configuration of individual pixels, which produces a high frequency spatial pixel noise that is attenuated by viewing from sufficiently far away (Harmon & Julesz, 1973). In low frequency bands, pixel-boundary noise is not a problem because the spatial filtering insures that adjacent pixels vary only slightly in intensity. We explored the hypothesis of pixel-boundary noise with subject CJD, who showed a distance effect in band 5. At an intermediate viewing distance of 1.21 m, CJD squinted her eyes while viewing stimuli from band 5. By blurring the retinal image of the display in this way, performance improved approximately to the level of the furthest viewing distance.

To summarize, the only significant effect of distance that we observed was a lowering of performance at near viewing distances relative to the furthest distance. This impairment occurred primarily in bands 4 and 5. In these bands, the spatial quantization of the display (90 × 90 square-shaped pixels) produces artifactual high spatial frequencies that mask the target. These artifactually produced spatial frequencies can be attenuated by deliberate blurring (squinting), or by producing displays with higher spatial resolution, or by increasing the viewing distance to the point where the pixel boundaries are attenuated by the optics of the eye and neural components of the visual modulation transfer function. In all cases, blurring improves performance and eliminates the slightly deleterious effect of a too small viewing distance. Thus, for correctly constructed stimuli, in the frequency ranges studied, there would be no significant effect of viewing distance on performance. This finding is in agreement with the results of Legge et al. (1985), who examined reading rate rather than letter recognition. It is in stark disagreement with the results of sinewave detection experiments in which retinal frequency is critical—see Sperling (1989) for an explanation.

DISCUSSION

A comparison of performance in different frequency bands shows that subjects perform better the higher the frequency band; and subjects require the smallest signal-to-noise ratio in the highest frequency band. To determine whether performance in high frequency bands is good because humans are more efficient in utilizing high-frequency information, or because there is objectively more information in the high-frequency images, or both, requires an investigation of the performance of an ideal observer. The performance of the ideal observer is the measure of the objective presence of information. Human performance results from the joint effect of the objective presence of information and the ability of humans to utilize that information. Human efficiency is the ratio of human performance to ideal performance.

Ideal discriminator

Definition. An ideal discriminator makes the best possible decision given the available data and the interpretation of "best." The performance of the ideal discriminator defines the objective utility of the information in the stimulus. We prefer the name *ideal discriminator*, rather than *ideal observer*, because it indicates the critical aspect of performance under consideration, but we occasionally use *ideal observer* to emphasize the relations to a large, relevant literature on this subject. Our purposes in this section are first, to derive an ideal discriminator for the letter identification task, second, to develop a practical working approximation to this discriminator, and third, to compare the performance of the human with the ideal discriminator.

Although ideal observers have recently come into greater use in vision research, the applications have focused primarily on determining the limits of performance for relatively low-level visual phenomena. For example, Barlow (1978, 1980), and Barlow and Reeves (1979) investigated the perception of density and of mirror symmetry; Geisler (1984) investigated the limits of acuity and hyperacuity; Legge, Kersten and Burgess (1987) examined the pedestal effect; Kersten (1984) studied the detection of noise patterns; and Pelli (1981) detailed the roles of internal visual noise. Geisler (1989) provides an overview of efficiency computations in early vision. Our application differs from these in that we expand the techniques and apply them to a higher perceptual/cognitive function, letter recognition.

For the letter identification task, the ideal discriminator is conceptually easy to define. A particular observed stimulus, x, representing an unknown letter plus noise, consists of an intensity value (one of 256 possible values) at each of 90 × 90 locations. The discriminator's task is to make the correct choice as frequently as possible from among the 26 alternative letters.

The likelihood of observing stimulus x, given each of the 26 possible signal alternatives, can be computed when the probability density function of the added noise is known exactly. The optimal decision chooses the letter that has the highest likelihood of yielding x. The expected performance of the ideal discriminator is computed by summing its probability of a correct response over the 256^{8100} possible stimuli (256 gray levels, 90 × 90 pixels). Unfortunately,

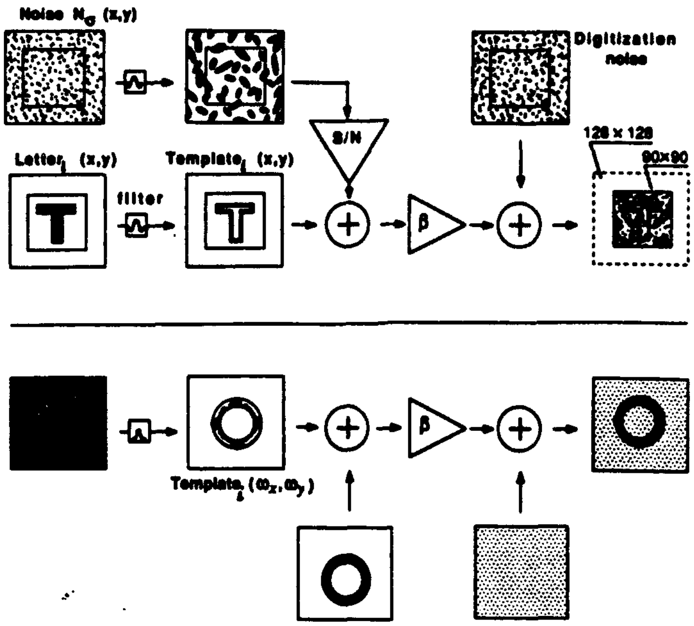


Fig. 6. Flow chart of the experimental procedures that are modelled by the ideal discriminator analysis. Upper half indicates space-domain operations; lower half indicates the corresponding operations in the frequency domain. Computations are carried out on 128 × 128 arrays; the subject sees only the center 90 × 90 pixels. A random letter and a random noise field are each filtered by the same filter (b); the noise is amplified to provide the desired signal-to-noise ratio; the letter and noise are added, the output is scaled and quantized (represented by the addition of digitization noise), and the result is shown to the subject. In the frequency domain ω_x, ω_y, the bandpass filter selects an annulus, whereas the quantization noise is uniform over ω_x, ω_y.

when there is both bandpass filtered and intensity quantization, the usual simplifications that make this enormous computation tractable are not applicable.

As an alternative to computing the expected performance of the ideal discriminator, one can compute its performance with a particular subset of the possible stimuli—the stimuli that the subject actually viewed or, preferably, a larger set of stimuli for more reliable estimation. This Monte Carlo simulation of the performance of the ideal discriminator is a tractable computation that yields an estimate of expected performance.

Derivation. Stimulus construction is diagrammed in Fig. 6 which shows the equivalent operations in the space and the frequency domains. To derive an ideal discriminator, we need to carefully review the processes of stimulus construction. We use uppercase letters to represent quantities in the frequency domain and lowercase letters to represent quantities in the space domain. A letter is defined by a 90 × 90 array that takes the value 1 at the letter locations and 0 at the background locations. When this array is spatially filtered in band b, it defines the *letter template* $t_{i,b}(x,y)$, where i indicates the particular letter, b the frequency band, and x, y the pixel location. We write $T_{i,b}(\omega_x, \omega_y)$ for the Fourier series coefficient of $t_{i,b}$ indexed by frequency.

An unknown stimulus $u_{i,b}(x,y)$ to be viewed by a subject is produced by adding filtered $n_b(x,y)$ with post-filtering variance σ_N^2, to the template $t_{i,b}(x,y)$, where letter identity i is unknown to the subject. The stimulus is scaled and digitized (quantized) to 256 levels prior to presentation, contributing an additional source of noise $q_{i,b}(x,y)$, called digitization noise. Finally, a d.c. component (dc) is added to $u_{i,b}$ to bring the mean luminance level to 128. These steps are diagrammed in Fig. 6 which shows both the space-domain and the corresponding frequency-domain operations. The space-domain computation is encapsulated in equations (3):

$$u_{i,b}(x,y) = \beta_{i,b}[t_{i,b}(x,y) + n_b(x,y)] \qquad (3a)$$

$$u_{i,b}(x,y) = \beta_{i,b}[t_{i,b}(x,y) + n_b(x,y)] + q_{i,b}(x,y) + dc. \qquad (3b)$$

The scaling constant $\beta_{i,b}$, limits the range of real values for each pixel, prior to quantization, to [−0.5, 255.5]. The degree of scaling is determined by the maximum and minimum values in the function $t_{i,b}+n_b$. Note that the extreme values in the image are determined by σ_{N_2} which is adjusted to yield the appropriate s/n for each condition; the values of $t_{i,b}$ are fixed prior to scaling. Specifically:

$$\beta_{i,b}=\frac{256}{\max(t_{i,b}+n_b)-\min(t_{i,b}+n_b)}. \quad (4)$$

As a result of bandpass filtering, the noise samples in adjacent pixels are strongly dependent on each other. Therefore, the discriminator problem is best approached in the Fourier domain, where the random variables $\{N_b(\omega_x,\omega_y)\}$ are jointly independent because the filtering operations simply scale the different frequency components without introducing any correlations (van Tress, 1968). The task of the ideal discriminator is to pick the template $t_{i,b}$ that maximizes the likelihood of $u_{i,b}$ with *a priori* knowledge of: (i) the fixed functions $t_{i,b}$, and their probabilities; and (ii) the densities of the jointly independent random variables $\{N_b(\omega_x,\omega_y)\}$. As is clear, $\beta_{i,b}$, σ_N^2, $\{Q_{i,b}(\omega_x,\omega_y)\}$, and $\{N_{i,b}(\omega_x,\omega_y)\}$ are all jointly distributed random variables characterized by some density f. To compute the likelihood of $u_{i,b}$ the ideal discriminator must integrate f over all possible values that may be assumed by the set of jointly distributed random variables, whose values are constrained only in that they result in a possible stimulus $u_{i,b}$. Unfortunately, no closed-form solution to this problem is available, forcing us to look for an alternative approach.

Bracketing. To estimate the performance of the ideal discriminator, we look for a tractable super-ideal discriminator that is better than the ideal but which is solvable. Similarly, we look for a tractable sub-ideal discriminator that is worse than the ideal. The ideal discriminator must lie between these two discriminators; that is, we bracket its performance between that of a "super-ideal" and a "sub-ideal" discriminator. The more similar the performance of the super- and sub-ideal discriminators, the more constrained is the ideal performance which lies between them.

Our super-ideal discriminator is told, *a priori*, the extact values for $\beta_{i,b}$ and σ_N^2 for each stimulus presentation. Therefore, it is expected to perform slightly better than the ideal discriminator which must estimate these values from the data. The sub-ideal discriminator estimates these same parameters from the presented stimulus in a simple but nonideal way. Therefore, it is expected to perform slightly worse than the ideal discriminator. The computational forms used to compute $\beta_{i,b}$ and σ_N^2 for the sub-ideal discriminator are presented in the Appendix, along with the derivation of the likelihood estimator used by both discriminators. A complete discussion of these derivations and the problems associated with the formulation of an ideal discriminator for such complex stimuli is presented in Chubb, Sperling and Parish (1987).

Performance of the bracketed discriminator. The super- and sub-ideal discriminators were tested in a Monte Carlo series of trials, in which they each were confronted with 90 stimuli in each of the frequency bands at each of seven s/n values chosen to best estimate their 50% performance point. The s/n necessary for 50% correct discriminations was estimated by an inverse interpolation of the best fitting logistic function. The derived $(s/n)_{50\%}$ is the measure of performance of a discriminator. The mean ratio, across frequency bands, of

$$(s/n)_{50\%}\ \text{sub-ideal}/(s/n)_{50\%}\ \text{super-ideal}$$

is about 2 (approx. 0.3 $\log_{10}$ units). The ratio does not depend on the criterion of performance.

Efficiency of human discrimination

In all conditions, human subjects perform worse than the sub-ideal discriminator. Notably, with no added luminance noise, the subideal (and, of course, the ideal) discriminator function perfectly, even in b_0 where subject performance is at chance, and in b_1 where subjects reached asymptote at about 50% correct.

Data from the subjects are plotted with the $(s/n)_{50\%}$ sub-ideal and $(s/n)_{50\%}$ super-ideal in Fig. 5. For comparison, Fig. 5 also shows the performance of a correlator discriminator which chooses the letter template that correlates most highly with the stimulus in the space domain. In the coordinates of Fig. 5 ($\log_{10}s/n$ vs $\log_{10}f$ where f represents the mean 2D spatial frequency of the band), the vertical distance d from the human data $\log(s/n)_{50\%}$, *human* down to the bracketed discriminator $\log(s/n)_{50\%}$, *ideal* represents the $\log_{10}$ of the factor by which the bracketed discriminator outperforms the human observer at that value of f. For the purpose of specifying efficiency, we assume the ideal discriminator lies at the mid-point of the sub and super-ideal discriminators in Fig. 5. The

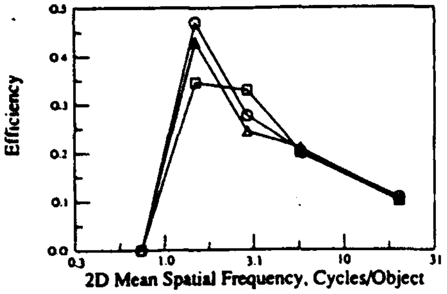


Fig. 7. Discrimination efficiency as a function of the mean frequency of a 2-octave band (in cycles per letter height) indicated on a logarithmic scale. Data are shown for three observers: △ = SAW, □ = RS, ○ = DHP. The viewing distance is 2.21 m, which is representative of all viewing distances tested.

efficiency *eff* of human discrimination relative to the bracketed discriminator is $eff = 10^{-2d}$, where:

$$d = \log(s/n)_{50\%,\,human} - \log(s/n)_{50\%,\,ideal}.$$

The values of *eff* in each object frequency band are shown in Fig. 7. In band 0, *eff* is zero because human performance never reaches 50%; indeed, it never rises significantly above 4% (chance). In band 1, human performance asymptotically climbs close to 50% as s/n approaches infinity; $eff \approx 0$. In band 2, *eff* reaches its maximum of 35–47% (depending on the subject), and it declines rapidly with increasing frequency (b_3–b_5).

The 42% average efficiency in band 2 is similar in magnitude to the highest efficiencies observed in comparable studies. For example, efficiency has been determined for detecting various kinds of patterns in arrays of random dots (Barlow, 1978, 1980; van Meeteren & Barlow, 1981), tasks which, like ours, may require significantly cognitive processing. In a wide range of conditions, the highest efficiencies observed were about 50%, and frequently lower. Van Meeteren and Barlow (1981) also found that efficiency was perfectly correlated with object spatial frequency and was independent of retinal spatial frequency.

Spatial correlator discriminator. A correlator discriminator cross-correlates the presented stimulus with its memory templates and chooses the template with the highest correlation. Correlation can be carried out in the space or in the frequency domain. Correlation is an efficient strategy when noise in adjacent pixels is independent and when members of the set of signals have the same energy; both of these conditions are violated by our stimuli. However, when sufficient prior information is available to subjects, they do appear to employ a cross-correlation strategy (Burgess, 1985).

It is interesting to note that the performance of the spatial correlator discriminator over the middle range of spatial frequencies is quite close to the performance of the sub-ideal discriminator. At high spatial frequencies, correlator performance degenerates, due to its inability to focus spatially on those pixel locations that contain the most information. A spatial correlator that optimally weighted spatial locations, could overcome the spatial focusing problem at high frequencies. (Spatial focusing is treated in the next section.)

At all frequencies, the spatial correlator is nonideal because noise at spatial adjacent pixels is not independent. At low spatial frequencies, the nonindependence of adjacent locations becomes extreme and the correlator fails miserably. This points out that, for our stimuli, correlation detection is better carried out in the frequency domain because there the noise at different frequencies is independent. The qualitative similarity between the correlator discriminator and the subjects' data suggests that the subjects might be employing a spatial correlation strategy, augmented by location weighting at high frequencies.

Lowest spatial frequencies sufficient for letter discrimination. Band 2 corresponds to a 2-octave band with a peak frequency of 1.05 c/object (vertical height of letters) and a 2D mean frequency of 1.49 c/object. At the four viewing distances, 1.05 c/object corresponds to retinal frequencies of 0.074, 0.234, 0.739 and 2.34 c/deg of visual angle. We observe perfect scale invariance: all of these retinal frequencies, and hence the visual channels that process this information, are equally effective in achieving the high efficiency of discrimination.

The finding that b_2 with a center frequency of 1.05 c/object and a $\frac{1}{2}$ amplitude cutoff at 2.1 c/object is critical for letter discrimination is in good agreement with previous findings of both Ginsburg (1978) for letter recognition and Legge et al. (1985) for reading rate. Legge et al. used low-pass filtered stimuli, which included not only spatial frequencies within an octave of 1 c/object (b_2) but also included all lower frequencies. From the present study, we expect human performance with low-pass and with band-pass spatial filtering to be quite similar up to 1 c/object because the lowest frequency bands, when presented in isolation, are perceptually useless (at least when presented alone).

It is an important fact that our subjects actually performed better, in the sense of achieving criterion performance at a lower s/n ratio, at higher frequency bands than b_2. This is explained by the increase in stimulus information in higher frequency stimuli. Increased information more than compensates for the subjects' loss in efficiency as spatial frequency increases.

Components of discrimination performance

Though the performance of the bracketed ideal discriminator is useful in quantifying the informational utility of the various bands, it is instructive to consider the changing physical structure of the stimuli as well. What components of the stimuli actually lead to a gain in information with increasing frequency? According to Shannon's theorem (Shannon & Weaver, 1949), an absolutely bandlimited 1-D signal can be represented by a number of samples m that is proportional to its bandwidth. When the signal-to-noise ratio in each sample s_i/n_i is the same, the overall signal-to-noise ratio s/n grows as $\sqrt{m}$. In the space domain, our filters were constructed (approximately) to differ only in scale but not in the shape of their impulse responses. Therefore, when the mean frequency of a filter band increased by a factor of 2, the bandwidth also increased by 2. Since the stimuli are 2D, the effective number of samples increases with the square of frequency, and the increase in effective s/n ratio is proportional to m. This expected improvement with frequency, based simply on the increase in effective number of samples, is indicated by the oblique parallel lines of Fig. 5 with slope of -1. The expected improvement in threshold s/n due simply to the linearly increasing bandwidth of the bands does a reasonable job of accounting for the improvement in performance for both human and bracketed discriminators between b_2 and b_5.

Performance of all discriminators improves faster with frequency between 0.39 and 1.5 c/object and between 5.8 and 22 c/object than is predicted from the bandwidths of the images. A slope steeper than -1 means that there is more information for discriminating letters in higher frequency bands even when the number of independent samples is kept the same in each band. Once sampling density is controlled, just how much information letters happen to contain in each frequency band is an ecological property of upper-case letters.

Increasing spatial localization with increasing frequency band. From the human observer's point of view, the letter information in low-pass filtered images is spread out over a large portion of the total image array. In high spatial-frequency images, the letter information is concentrated in a small proportion of the total number of pixels. In high spatial-frequency images, a human observer who knows which pixels to attend will experience an effective s/n that is higher than an observer who attends equally to all pixels. In this respect, humans differ from an ideal discriminator. The ideal discriminator has unlimited memory and processing resources, does not explicitly incorporate any selective mechanism into its decision, and uses the same algorithm in all frequency bands. Information from irrelevant pixels is enmeshed in the computation but cancels out perfectly in the letter-decision process. To understand human performance, however, it is useful to examine how, with our size-scaled spatial filters, letter information comes to be occupy a smaller and smaller fraction of the image array as spatial frequency increases.

Here we consider three formulations of the change in the internal structure of the images with increasing spatial frequency: (1) spatial localization; (2) correlation between signals; and (3) nearest neighbor analysis. We have already noted that, in our images, the information-rich pixels become a smaller fraction of the total pixels as frequency band increases. Indeed, this reduction can be estimated by computing the information transmitted at any particular pixel location or, more appropriately for estimating noise resistance, by computing the variance of intensity (at that pixel location) over the set of 26 alternative signals.

To demonstrate the degree of increasing localization with increasing frequency, the variance (over the set of 26 letter templates) was computed at each pixel location (x, y). *Total power*, the total variance, is obtained by summing over pixel locations. The number of pixel locations needed to achieve a specific fraction of the total power is given in Fig. 8, with frequency band as a parameter. These curves describe the spatial distribution of information in the latter templates. If all pixels were equally informative, exactly half of the total number of pixels would be needed to account for 50% of the total power. The solid curves in Fig. 8 show that the number of pixels needed to convey any percentage of total signal power, decreases as the

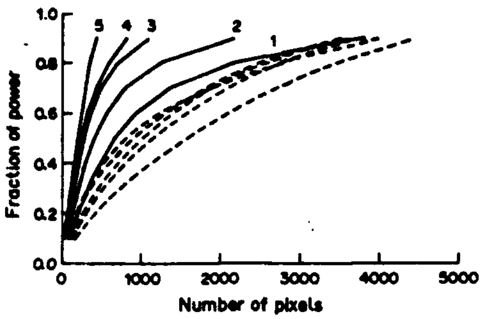


Fig. 8. Fraction of total power contained in the n most extreme-valued pixels as a function of n (out of 8100). Solid lines indicate the power fractions for signals; the curve parameter indicates the filter band. Dashed lines indicate power fractions for filtered noise fields. Although power fractions from successive bands of noise are too close to label, they generally fall in the same left–right 5–0 order as those for signal bands.

Table 3. Average pairwise correlations and nearest neighbors (Euclidean distance × 10^{-3})

Band	Correlations	Nearest neighbor
0	0.94	0.01
1	0.91	0.30
2	0.58	1.2
3	0.38	2.3
4	0.33	3.1
5	0.31	4.1

frequency band increases. These information distribution curves are an ecological property of our set of letter stimuli; different curves would be needed describe other stimulus sets.

The dashed curves in Fig. 8 were derived from random noise filtered in each of the six frequency bands (b_0–b_5). The distribution of noise power is very similar between the various bands, enormously more so than the distribution of signal power. For our letter stimuli, stimulus information coalesces to a smaller number of spatial locations as spatial frequency increases.

Correlation between signals. A more abstract way of describing the change of information with bandwidth is to note that letters become less confusible with each other in the higher frequency bands. A good measure of confusibility is the average pairwise correlation between the 26 letter templates in each frequency band (Table 3). The average correlation between letter templates diminishes from 0.94 in band 0 to 0.31 in band 5. In a band in which templates have a pairwise correlation over 0.9, the overwhelming amount of intensity variation ("information") is useless for discrimination. Small wonder that subjects fail completely in this band. Overall, performance of the ideal discriminator and of observers improves as the correlation decreases, but there is no obvious way to use the pairwise correlation between templates to predict performance.

Nearest neighbors. The analysis of nearest neighbors is a useful technique for predicting accuracy by the analysis of the possible causes of errors. We can regard a filtered image t_i of letter i as a vector in a space of dimensionality 8100 (90 × 90 pixels). When noise is added, the possible positions of t_i are described by a cloud whose dimensions are determined by the s/n ratio. A neighboring letter k may be confused with letter i when the cloud around t_i envelopes t_k. The closer the neighbor, the greater the opportunity for error. Table 3 gives the average normalized distance to the nearest neighbor in each of the bands. The increase in distance to the nearest neighbor reflects the improvement in the representation of signals as spatial frequency increases.

We consider possible causes of lower efficiency of discrimination in bands below b_2. The letters in these bands have high pair-wise correlations and the mean band frequency is less than the object frequency. This means that letters differ only in subtle differences of shading, a feature that we usually do not think of as shape. Observers would need to be able to utilize small intensity differences to distinguish between letters. To eliminate an alternative explanation (the smaller number of frequency components in the low-frequency bands), we conducted an informal experiment with a lower fundamental frequency. The fundamental frequency, which is outside the band, nevertheless determines the spacing of frequency components within the band. Reducing the fundamental frequency of the letter by one-half increases the number of frequency components in the band by a factor of 4. (A 256 × 256 sampling grid was used rather than 128 × 128.) These 4 × more highly sampled stimuli were not more discriminable than the original stimuli. This suggests that the internal letter representation (template) that subjects bring with them to the experiment cannot utilize low-frequency information, even when it is abundantly available. Whether, with sufficient training, subjects could learn to use low spatial frequencies to make letter discriminations is an open question.

SUMMARY AND CONCLUSIONS

1. Visual discrimination of letters in noise, spatially filtered in 2-octave wide bands, is

independent of viewing distance (retinal frequency) but improves as spatial frequency increases.

2. The improvement in performance with increasing spatial frequency results mainly from an increase in the objective amount of information transmitted by the filters with increasing frequency (because filter bandwidth was proportional to center frequency) which is manifested as objectively less confusible stimuli in the higher bands.

3. The comparison of human performance with that of an estimated ideal discriminator demonstrates that humans achieve optimal discrimination (a remarkable 42% efficiency) when letters are defined by a 2-octave band of spatial frequencies centered at 1 cycle per letter height (mean frequency 1.5 c/letter). This high efficiency of discrimination is maintained over a 32:1 range of viewing distances.

4. Detection efficiency was invariant over a range of retinal spatial frequencies in which the contrast threshold for detection of sine gratings (the modulation transfer function, MTF) varies enormously. The independence of detection performance and retinal size held for all frequency bands.

5. A part of the loss of human efficiency in discrimination as spatial frequency exceeded 1 c/object height may have been due to the subjects' inability to identify, to selectively attend, and to utilize the smaller fraction of information-rich pixels in the higher frequency images.

6. Finally, it is important to note that without the comparison to the ideal observer, we would not have been able to understand the components of human performance in the different frequency bands.

Acknowledgements—We acknowledge the large contribution of Charles Chubb to the formulation and solution of the ideal discriminator. We thank Michael S. Landy for helpful comments and Robert Picardi for skillful technical assistance. The project was supported by USAF, Life Sciences Directorate, Visual Information Processing Program, grants 85-0364 and 88-0140.

REFERENCES

Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Research, 18,* 637–650.

Barlow, H. M. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society, London B, 290,* 71–82.

Barlow, H. B. & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research, 19,* 783–793.

Burgess, A. (1985). Visual signal detection—III. On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America A, 2(9),* 1498–1507.

Burgess, A. (1986). Induced internal noise in visual decision tasks. *Journal of the Optical Society of America A, 3,* 93.

Burt, P. J. & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications, Com-34(4),* 532–540.

Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, London 197,* 551–566.

Carlson, C. R., Moeller, J. R. & Anderson, C. H. (1984). Visual illusions without low spatial frequencies. *Vision Research, 24,* 1407–1413.

Chubb, C., Sperling, G. & Parish, D. H. (1987). Designing psychophysical discrimination tasks for which ideal performance is computationally tractable. Unpublished manuscript, New York University, Human Information Processing Laboratory.

Davidson, M. L. (1968). Perturbation approach to spatial brightness interaction in human vision. *Journal of the Optical Society of America A, 58,* 1300–1309.

Fiorentini, A., Maffei, L. & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception, 12,* 195–201.

Geisler, W. S. (1984). Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A, 1,* 775–782.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review, 21,* 267–314.

Ginsburg, A. P. (1971). Psychological correlates of a model of the human visual system. In *Proceedings of the National Aerospace Electronics Conference (NAECON)* (pp. 283–290). Ohio: IEEE Trans. Aerospace Electronic Systems.

Ginsburg, A. P. (1978). Visual information processing based on spatial filters constrained by biological data. Aerospace Medical Research Laboratory, 1(2), Dayton, Ohio.

Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display designs: An explanation of how we see certain objects. *Proceedings of SID, 21,* 219–227.

Ginsberg, A. P. & Evans, P. W. (1979). Predicting visual illusions from filtered imaged based on biological data. *Journal of the Optical Society of America A, 69,* 1443.

Harmon, L. D. & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science, 180,* 1194–1197.

Janez, L. (1984). Visual grouping without low spatial frequencies. *Vision Research, 24,* 271–274.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Research, 24,* 1977–1990.

Legge, G. E., Pelli, D. G., Rubin, G. S. & Schleske, M. M. (1985). Psychophysics of reading—I. Normal vision. *Vision Research, 25(2),* 239–252.

Legge, G. E., Kersten, D. & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A, 4(2),* 391–404.

van Meeteren, A. & Barlow, H. B. (1981). The statistical efficiency for detecting sinusoidal modulation of average dot density in random figures. *Vision Research, 21,* 765–777.

van Nes, F. L. & Bouman, M. A. (1967). Spatial modulation transfer in the human eye. *Journal of the Optical Society of America, 57,* 401–406.

Norman, J. & Ehrlich, S. (1987). Spatial frequency filtering and target identification. *Vision Research, 27(1)*, 97–96.

Parish, D. H. & Sperling, G. (1987a). Object spatial frequencies, retinal spatial frequencies, and the efficiency of letter discrimination. Mathematical Studies in Perception and Cognition, 87–8. New York University, Department of Psychology.

Parish, D. H. & Sperling, G. (1987b). Object spatial frequency, not retinal spatial frequency, determines identification efficiency. *Investigative Ophthalmology and Visual Science (ARVO Suppl.), 28(3)*, 359.

Pavel, M., Sperling, G., Riedl, T. & Vanderbeek, A. (1987). The limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American sign language. *Journal of the Optical Society of America A, 4*, 2355–2365.

Pelli, D. G. (1981). Effects of visual noise. Ph.D. dissertation, University of Cambridge, England.

Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Sperling, G. (1989). Three stages and two systems of visual processing. *Spatial Vision, 4 (Prazdny Memorial Issue)*, 183–207.

Sperling, G. & Parish, D. H. (1985). Forest-in-the-Trees illusions. *Investigative Ophthalmology and Visual Science (ARVO Suppl.), 26*, 285.

Tanner, W. P. & Birdsall, T. G. (1958). Definitions of d' and n as psychophysical measures. *Journal of the Acoustical Society of America, 30*, 922–928.

van Tress, H. L. (1968). *Detection, estimation and modulation theory*. New York: Wiley.

Winer, B. J. (1971). *Statistical principles in experimental psychology*. New York: McGraw-Hill.

APPENDIX

Both sub-ideal and super-ideal discriminators must compute estimates of the likelihood that the stimulus $u_{k,b}$ was produced with template $t_{i,b}$ and noise n_b, where k is the letter used to generate the stimulus, i is an arbitrary letter, and b indexes spatial frequency band. Let x be an index on the pixels of the image: $1 \leqslant x \leqslant 8100$, for the 90×90 images of the experiments.

For the Monte Carlo simulations of the super-ideal discriminator, the unknown stimulus parameters, $\alpha_{i,b}$ and σ_N^2 are computed during stimulus construction, and their exact values are supplied to the discriminator *a priori*. The sub-ideal discriminator, however, must estimate these parameters from the data as follows.

Sub-Ideal Parameter Estimation

Recall that stimulus contrast is modulated for any pixel x in the image:

$$u_{k,b}[x] = \beta_{i,b}[t_{i,b}(x) + n_b(x)] + q_{i,b}(x). \tag{A1}$$

The scaling constant $\beta_{i,b}$ limits range of real values for each pixel, prior to quantization, to the open interval $(-0.5, 255.5)$; the addition of $q_{i,b}[x]$, called quantization noise, rounds off pixel values to integers.

For each bandpass filtered template $t_{i,b}$, we first compute the correlation $\rho_{k,i}$ of the template to the stimulus $u_{k,b}$:

$$\rho_{k,i} = \frac{\sum_x u_{k,b}(x)\, t_{i,b}(x)}{\left\{\sum_x [u_{k,b}(x)]^2\right\}^{1/2} \left\{\sum_x [t_{i,b}(x)]^2\right\}^{1/2}}. \tag{A2}$$

To compute the likelihood estimates for each template $t_{i,b}$, we must be able to reverse the effect of $\beta_{i,b}$. Thus we define $\alpha_{i,b} = 1/\beta_{i,b}$ and choose $\alpha_{i,b}$ so as to minimize the expression:

$$\sum_x [\alpha_{i,b} u_{k,b}(x)]^2 = \sum_x [\rho_{k,i} t_{i,b}(x)]^2. \tag{A3}$$

Solving for $\alpha_{i,b}$ gives us:

$$\alpha_{i,b} = \rho_{k,i} \left\{ \frac{\sum_x [t_{i,b}(x)]^2}{\sum_x [u_{k,b}(x)]^2} \right\}^{1/2}. \tag{A4}$$

Finally we set:

$$\sigma_N^2 = \frac{1}{X} \sum_{x=1}^{X} [\alpha_{i,b} u_{k,b}(x) - t_{i,b}(x)]^2 \tag{A5}$$

where $X = 8100$, the number of pixels in the image.

Likelihood Estimation

With estimates of σ_N^2 and $\alpha_{i,b}$ for the sub-ideal discriminator, and the *a priori* values for the super-ideal discriminator, we can formulate a maximum likelihood estimator. By rearranging terms of equation (A1) and dividing both sides by β yields:

$$\frac{u_{k,b}(x)}{\beta} - t_{i,b}(x) = n_b(x) + \frac{q_{i,b}(x)}{\beta}. \tag{A6}$$

Substituting $\alpha_{i,b}$ for $1/\beta$, and by transposing into the frequency domain, denoted by upper-case letters and indexed by ω, we have:

$$\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega) = N_b(\omega) + \alpha_{i,b} Q_{i,b}(\omega). \tag{A7}$$

Note that the left side of equation (A7) is simply a difference image between the stimulus $U_{k,b}(\omega)$ and the template $T_{i,b}(\omega)$. This difference is exactly equal to the sum of the luminance and quantization noise only when the correct template is chosen ($i = k$). When the incorrect template is chosen ($i \neq k$) the right hand side of equation (A7) is equal to the sum of the noise sources plus some residue that is equal to $T_{k,b}(\omega) - T_{i,b}(\omega)$. Under the assumption that quantization noise can be modeled as independent additive noise in the frequency domain, the density Λ of the joint realization of the right-hand side of equation (A7) is given by:

$$\Lambda = \prod_\omega \frac{X}{\pi[\sigma_Q^2 \alpha_{i,b}^2 + \sigma_N^2 |F_b(\omega)|]^2} \times \exp\left[\frac{-x\,|\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega)|^2}{\alpha_{i,b}^2 \sigma_Q^2 + \sigma_N^2 |F_b(\omega)|^2}\right] \tag{A8}$$

where $F_b(\omega)$ is simply the kernel of filter b, in the frequency domain. Dropping the multiplicative term in equation (A8), which does not depend on the template T, and taking logs, the ideal discriminator chooses the template that minimizes:

$$\sum_\omega \frac{X\,|\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega)|^2}{\alpha_{i,b}^2 \sigma_Q^2 + \sigma_N^2 |F_b(\omega)|^2}. \tag{A9}$$

Finally, it is more convenient to compute the power of the quantization noise in the space domain (σ_q^2) than in the frequency domain (σ_Q^2); $\sigma_q^2 = \sigma_Q^2$. Spatial quantization noise, $q_{i,b}(x)$, is uniformly distributed on the interval $[-0.5, 0.5)$, so that σ_q^2 is computed as:

$$\int_{-0.5}^{0.5} x^2\, dx \tag{A10}$$

and is equal to 1/12.

Vision Res. Vol. 31, No. 5, pp. 859–876, 1991
Printed in Great Britain. All rights reserved

0042-6989/91 $3.00 + 0.00

THE KINETIC DEPTH EFFECT AND OPTIC FLOW—II. FIRST- AND SECOND-ORDER MOTION

MICHAEL S. LANDY,[1] BARBARA A. DOSHER,[2] GEORGE SPERLING[1] and MARK E. PERKINS[1]

[1]Psychology Department, New York University, NY 10003 and [2]Psychology Department, Columbia University, NY 10027, U.S.A.

(*Received 24 August 1989; in revised form 1 May 1990*)

Abstract—We use a difficult shape identification task to analyze how humans extract 3D surface structure from dynamic 2D stimuli—the kinetic depth effect (KDE). Stimuli composed of luminous tokens moving on a less luminous background yield accurate 3D shape identification regardless of the particular token used (either dots, lines, or disks). These displays stimulate both the 1st-order (Fourier-energy) motion detectors and 2nd-order (nonFourier) motion detectors. To determine which system supports KDE, we employ stimulus manipulations that weaken or distort 1st-order motion energy (e.g. frame-to-frame alternation of the contrast polarity of tokens) and manipulations that create *microbalanced* stimuli which have no useful 1st-order motion energy. All manipulations that impair 1st-order motion energy correspondingly impair 3D shape identification. In certain cases, 2nd-order motion could support limited KDE, but it was not robust and was of low spatial resolution. We conclude that 1st-order motion detectors are the primary input to the kinetic depth system. To determine minimal conditions for KDE, we use a two frame display. Under optimal conditions, KDE supports shape identification performance at 63–94% of full-rotation displays (where baseline is 5%). Increasing the amount of 3D rotation portrayed or introducing a blank inter-stimulus interval impairs performance. Together, our results confirm that the human KDE computation of surface shape uses a global optic flow computed primarily by 1st-order motion detectors with minor 2nd-order inputs. Accurate 3D shape identification requires only two views and therefore does not require knowledge of acceleration.

KDE   Kinetic depth effect   Structure from motion   Shape   Optic flow

INTRODUCTION

When a collection of randomly positioned dots moves on a CRT screen with motion paths that are projections of rigid 3D motion, a human viewer perceives a striking impression of three-dimensionality and depth. This phenomenon of depth computed from relative motion cues is known as the kinetic depth effect (KDE; Wallach & O'Connell, 1953).

What are the important cues that lead to a 3D percept from such a display? Is it motion, or are there other important cues? If it is motion, then what kind of motion detection system(s) are used to support the structure-from-motion computation? Is a computation of velocity sufficient, or are more elaborate measurements necessary, such as of acceleration? These are the questions that we address in this paper.

In a series of recent papers (Dosher, Landy & Sperling, 1989a, b; Sperling, Landy, Dosher & Perkins, 1989; Sperling, Dosher & Landy, 1990), we examined the cues necessary for subjects to perceive an accurate representation of a 3D surface portrayed using random dot displays. In each trial of a new shape identification task we devised, subjects view a random dot representation of one of a set of 53 3D shapes and identify the shape and rotation direction. Shape identity feedback optimizes the subject's ability to compute shape from each type of motion stimulus. For accurate performance, the task requires either a 3D percept or a subject strategy that uses 2D velocity information in a manner that is computationally equivalent to that required to solve for 3D shape (Sperling et al., 1989, 1990; see the discussion of expt 2, below).

We have shown that the only cue used for the perception of three-dimensionality in these displays is motion (Sperling et al., 1989, 1990). Further experiments determined that global optic flow is used rather than the position information for individual dots, since accuracy remains high when dot lifetimes are reduced to as little as two frames (Dosher et al., 1989b). In that paper, we concluded that the input to the KDE computation is an optic flow generated by a 1st-order motion detection mechanism, such as the Reichardt detector (Reichardt, 1957). Two manipulations that perturb 1st-order motion energy mechanisms—flicker and polarity alternation—also interfered with KDE (Dosher et al., 1989b). In polarity alternation, dots change over time from black to white to black on a gray background. When compared to dots that remain white, polarity alternation was equally or slightly more detectable in a detection task, was poorer but still well above chance in a discrimination of direction of motion task (computed, presumably, using tracking of the dots or using more elaborate, 2nd-order motion detection mechanisms) but was useless for tasks requiring KDE or motion segregation. These latter two tasks require the evaluation of velocity in a number of locations simultaneously (Sperling et al., 1989). Shape identification performance in a range of conditions was shown to be monotonic with a computed index of 1st-order net directional power in the stimuli (Dosher et al., 1989b). Hence, for sparse dot stimuli, KDE depends upon a simple spatio-temporal (1st-order) Fourier analysis of multiple local areas of the stimulus.

In this paper, we further examine and generalize the contributions of several types of motion detectors to the optic flow computations used by the structure-from-motion mechanism.

MOTION ANALYSIS MODELS AND THE KDE

1st-order motion analysis

To motivate the stimulus conditions studied here, we begin by summarizing models of early motion detection and analysis. Several recent motion detection models (van Santen & Sperling, 1984, 1985; Adelson & Bergen, 1985; Watson & Ahumada, 1985) share as a common antecedent the model proposed by Reichardt (1957). We refer to this class of models as 1st-order motion detectors. Below, 2nd-order mechanisms involving additional processing stages will be discussed. In the Reichardt detector, luminance is measured at two spatial locations *A* and *B*. The measurement at position *A* is delayed in time, and then cross-correlated over time with the measurement at position *B*, resulting in a "half-detector" sensitive to motion from position *A* to *B*. A second such "half-detector" sensitive to motion from *B* to *A* is set in opponency with the first, resulting in the full motion detector. van Santen and Sperling (1984, 1985) have investigated this model along with extensions involving voting rules for combining outputs of many detectors to enable predictions of psychophysical experiments, resulting in their Elaborated Reichardt Detector (ERD).

An alternative way of characterizing motion detection is in the frequency domain. A motion detector can be built of several linear spatio-temporal filters. Each filter is sensitive only to energy in two of the four quadrants in spatio-temporal Fourier space (ω_x, ω_t). In other words, the filters are not *separable*. Their receptive fields are oriented in space-time, and thus they are sensitive to motion in a particular direction and at a particular scale (Adelson & Bergen, 1985; Burr, Ross & Morrone, 1986; Watson & Ahumada, 1985). The Fourier "energy" (the squared output of a quadrature pair of filters) in each of two opposing motion directions is computed, and put in opponency. This "motion energy detector", proposed by Adelson and Bergen (1985), and the ERD differ in their construction and in the signals available at the subunit level, but are indistinguishable at their outputs (Adelson & Bergen, 1985; van Santen & Sperling, 1985).

The structure-from-motion computation relies upon the measurement of image velocities at several image locations. The KDE shape identification task that we use here can be solved by categorizing velocity at six spatial locations into three categories: leftward, approximately zero, and rightward (Sperling et al., 1989). Thus, in order to discriminate the 53 test shapes by KDE, motion detection must be followed by at least some rudimentary local velocity calculation.

In order to signal velocity, the outputs of more than one such 1st-order motion detector must be pooled. Speed may be computed by pooling only two detectors (a motion and a "static" detector, Adelson & Bergen, 1985). To signal motion direction, signals must be pooled across a variety of orientations (Watson & Ahumada, 1985). Finally, in order to solve the "aperture problem" for more complex stimuli (Burt & Sperling, 1981; Marr & Ullman, 1981), signals may be pooled over a variety of directions and perhaps scales (Heeger, 1987).

In the previous paper (Dosher et al., 1989b), shape identification performance was shown to relate directly to the quality of the signal available from 1st-order motion detection mechanisms. Each stimulus consisted of a large number of dots on a gray background representing a 2D projection of dots on the surface of a smooth 3D shape under rotary oscillation. In one condition (contrast polarity alternation), the dots were first brighter than the background ("white-on-gray"), then darker than the background ("black-on-gray"), then bright again, in successive frames. For a dense random dot field (50% black/50% white) under simple planar motion, polarity alternation causes a percept of motion opposite to the true direction of motion (the "reverse-phi phenomenon", Anstis & Rogers, 1975); reverse-phi is thought to reflect a spatio-temporal Fourier analysis of the stimulus, since contrast reversal reverses the direction of motion of the lowest-frequency Fourier components (van Santen & Sperling, 1984). With contrast reversal, the outputs of 1st-order motion detection mechanisms no longer simply signal the intended direction and velocity of motion. Contrast reversal stimuli do not yield a depth-from-motion percept (Dosher et al., 1989b). We take this as evidence that the KDE relies upon input from a 1st-order motion analysis.

2nd-order motion analysis

For the sparse random dot stimuli (Dosher et al., 1989b), contrast polarity alternation eliminated the perception of structure from motion. Nonetheless, subjects could judge the direction of patches of contrast polarity alternating dots undergoing simple translation. What kind of a motion detector might be used to correctly judge the motion of a translating, polarity-alternating dot? One simple possibility would be to first apply a luminance nonlinearity to the input stimulus. For example, if the input stimulus were full-wave rectified about the mean luminance, the polarity-alternating stimulus would be converted to the equivalent of rigid motion of a white dot on a gray background. Thus, a full-wave rectifier of contrast followed by a 1st-order analyzer (such as those discussed above) would be capable of analyzing such a motion stimulus correctly (Chubb & Sperling, 1988b, 1989a, b).

A motion detection system consisting of a contrast nonlinearity followed by a 1st-order detector is one example of a wide class of "2nd-order detection mechanisms", each of which consists of a linear filtering of the input (spatial and/or temporal), followed by a contrast nonlinearity, followed by a standard 1st-order motion detection mechanism. A number of results demonstrate the existence of both 1st- and 2nd-order motion mechanisms and show the contribution of both to the perception of planar motion (Anstis & Rogers, 1975; Chubb & Sperling, 1988b, 1989a, b; Lelkens & Koenderink, 1984; Ramachandran, Rao & Vidyasagar, 1973; Sperling, 1976).

Can both 1st- and 2nd-order motion mechanisms be used by the KDE system? The polarity-alternating dots did not yield an effective KDE percept of our 3D shapes. If one accepts the existence of both 1st- and 2nd-order motion mechanisms, why didn't the 2nd-order system support KDE? The KDE stimuli were relatively small (3.7 × 4.2 deg) and viewed foveally (eye movements were permitted throughout the 2 sec stimulus duration). Evidence from studies of planar motion suggests that both systems were available under these conditions (Chubb & Sperling, 1988b). For polarity alternation stimuli, the most salient low frequency components from the 1st-order system were in the wrong direction. We assume that the 2nd-order system yields a correct (if attenuated) analysis. Bad shape identification performance may have resulted either from the perturbed 1st-order analysis or because of competition between the 1st- and 2nd-order systems (which signaled opposite directions of motion in some frequency bands). Our evidence (Dosher et al., 1989b) demonstrated that 1st-order system input is the predominant input to KDE, but it did not exclude the possibility of input from 2nd-order motion detection mechanisms. To approach that question, we consider a KDE stimulus that produces a simple 2nd-order motion analysis, but to which the 1st-order motion system is, statistically, blind.

Microbalanced motion stimuli

Chubb and Sperling (1988b) defined a class of stimuli, called *microbalanced*, among which are stimuli with the properties that we desire. In expt 1 we concentrate on two examples of microbalanced motion stimuli. These stimuli are random in the sense that any given stimulus is a realization of a random process. As proven by Chubb and Sperling (1988b), if a stimulus is microbalanced then the expected output of every 1st-order detector (ERD or motion energy detector) will be zero. Thus, Chubb and Sperling defined a class of stimuli for which a consistent motion signal requires a 2nd-order motion analysis, and showed that the 2nd-order analysis predicted observers' percepts for several examples of the class.

The polarity alternation stimulus is not microbalanced; any given frequency band does show consistent motion, with the lowest spatial frequencies signalling motion in the wrong direction. This stimulus can be transformed into a microbalanced one as follows: for each dot, choose the contrast polarity randomly and independently for every frame. Any given 1st-order detector will be just as likely to signal rightward motion as it is to signal leftward motion since it will either see the same contrast polarity across any successive pair of frames or it will see contrast polarity alternate, with equal probability. One question we examine in this paper is whether the motion signal available from 2nd-order mechanisms can be used to compute 3D structure.

We present two experiments. In the first, we examine performance on a shape identification task for a variety of KDE stimuli. Several types of stimuli provide good 1st-order motion. Others are microbalanced and hence can only be analyzed by 2nd-order mechanisms. Still others offer good 1st-order motion, but involve camouflage similar to that available in some of the microbalanced conditions. We find that 1st-order motion is used, and that input from 2nd-order mechanisms may also be used but is not as robust. In a second experiment, we examine the residual shape percept from two-frame KDE stimuli in order to determine whether a single velocity field is a sufficient cue for shape identification or whether acceleration also is needed.

EXPERIMENT 1. POLARITY ALTERNATION, MICROBALANCE, AND CAMOUFLAGE

In the first experiment, a shape discrimination task is used with a variety of displays. First, in order to sensibly compare results to our previous work (Sperling et al., 1989; Dosher et al., 1989b), there are control conditions that are identical to those of our previous experiments (the "Motion without density cue, standard speed, standard intensity" and "Motion with polarity alternation, standard speed, standard intensity" conditions of the preceding paper). In addition to dots, randomly positioned disks and lines are also used here in order to examine the effects of the foreground token used to carry the motion. The disk and line tokens are larger than the single pixel dots, and hence have more contrast energy. They enable us to test whether our previous failure to find KDE with polarity alternation resulted from the low contrast energy in the stimulus. Two forms of microbalanced stimuli are used, allowing us to test KDE shape identification performance with stimuli to which 1st-order motion detectors are blind. Finally, we examine stimuli in which moving textured tokens are camouflaged by a similarly textured background.

Method

Subjects. There were three subjects in this experiment. One was an author, and the other two were graduate students naive to the purposes of this experiment. All had normal or corrected-to-normal vision. There were slight differences in the conditions for each of the three subjects. These will be pointed out below.

White-on-gray dot stimuli. First, we briefly describe the stimuli that consist of bright dots moving on a gray background representing a variety of 3D shapes. This description will be somewhat abbreviated, since the same stimuli have been used in previous studies and more complete descriptions are available (Sperling et al., 1989). The other stimuli used in the present study result from simple image processing transformations applied to the white-on-gray dot stimuli.

Stimuli were based upon a fixed vocabulary of simple shapes consisting of bumps and concavities on a flat ground. The 3D shapes varied in the number, position, and 2D extent of these bumps and concavities. The process of generating the stimuli is illustrated in Fig. 1.

The first step in creating a stimulus involves the specification of a 3D surface. For a square area with sides of length s, a circle with diameter $0.9s$ is centered, and three fixed points, labeled 1, 2 and 3, are specified. For a given shape, one of two such sets of points is used (the upward-pointing triangle or the downward-pointing triangle, labeled u and d, respectively). The shape is specified as having a depth of zero outside of the circle. For each of the three identified points, the depth may be either $+0.5s$, 0.0, or $-0.5s$, which are labeled as $+$, 0, and $-$, respectively. The depth values for the rest of the figure were interpolated by using a standard cubic spline to connect the three interior points with the zero depth surround. Thus, there are 54 ways to designate a shape: u vs d, and for each of three interior points, $+$ vs 0 vs $-$. We designate a shape by denoting the triangle used, followed by the depth designations of the three points in the order shown in Fig. 1A. For example, $u-+0$

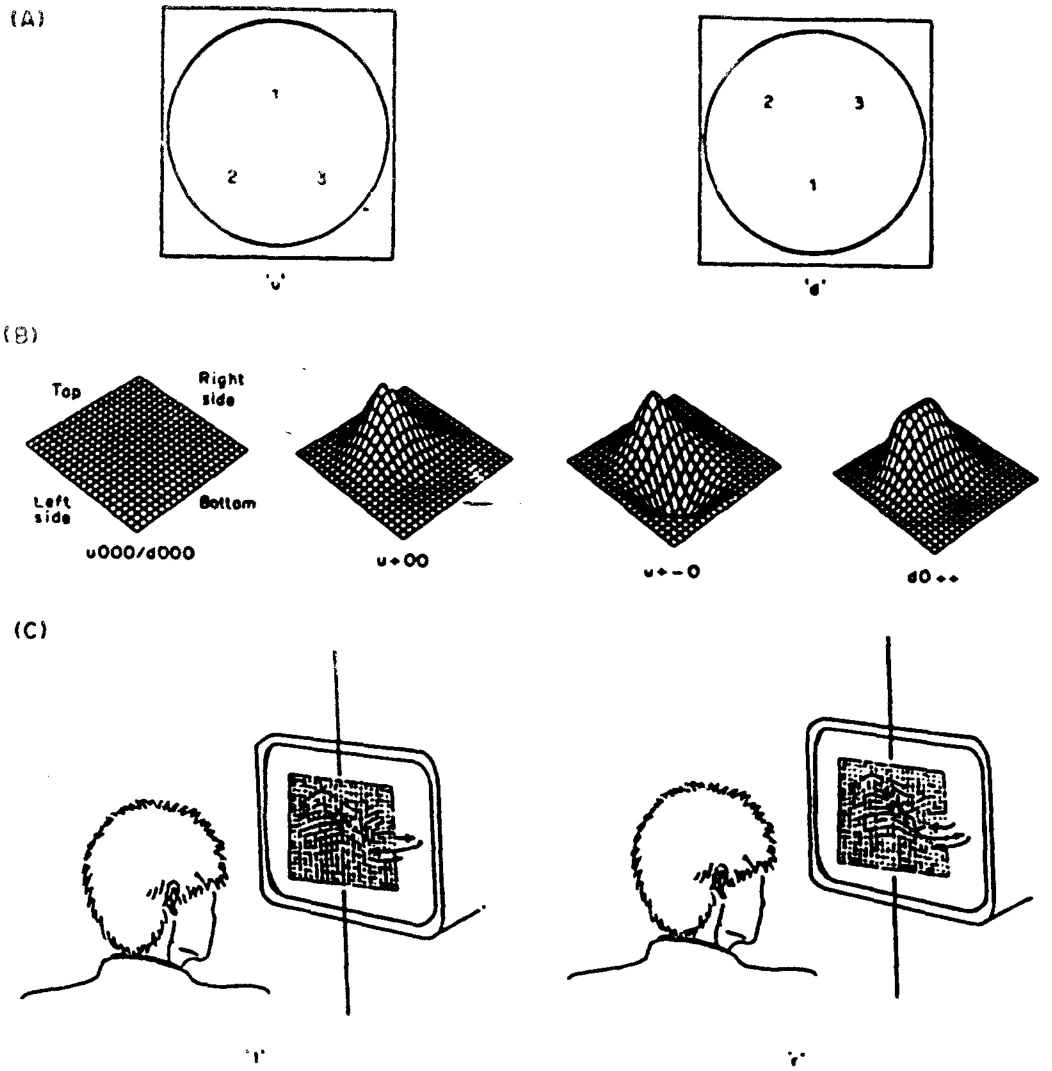


Fig. 1. Stimulus shapes, rotations, and their designations. (A) Shapes were constructed by choosing one of the two equilateral triangles represented here. Each point in the triangles was given a positive depth (i.e. toward the observer), zero depth, or negative depth, represented as +, 0 and −, respectively. A smooth shape splined these three points to zero depth values outside of the circle. A shape is designated by the choice of triangle (*u* or *d*), followed by the depth designations of the three points in the order given in the figure. (B) Some representative shapes generated by this procedure. All shapes consisted of a bump, concavity, or both, with a variation in position and extent of these areas. (C) Shapes were represented by a set of dots randomly painted on the surface of the shape, and wiggled about a vertical axis through the center of the display. The motion was a sinusoidal rotation that moved the object so as to face off to the observer's right, then his or her left, then back to face-forward (denoted *l*), or the reverse (denoted *r*).

is a shape with a bump in the upper-middle of the display, and a concavity in the lower-left (Fig. 1B). There are 53 distinct shapes, because *u*000 and *d*000 both denote a flat square.

Displays were generated by sprinkling dots randomly on the 3D surface generated by the spline, rotating that surface, and projecting the resulting dot positions onto the image-plane using parallel perspective. A large number of dots are chosen uniformly over a 2D area somewhat larger than the *s* by *s* square, and each dot's depth is determined by the cubic spline interpolant (where the zero depth of the surround is continued outside the square). This collection of dots is rotated about a vertical axis that is at zero depth and centered in the display. The rotation angle $\theta(k)$ is a sinusoidal "wiggle": $\theta(k) = \pm 25 \sin(2\pi k/30)$ deg, where k is the frame number within the 30 frame display. Thus, the display either rotated 25 deg to the right, then reversed its direction until it faced 25 deg to the left, then reversed its direction until it was again facing forward (labeled *l*), or rotated in the opposite manner (labeled *r*, see Fig. 1C). The displays presented these 3D collections of dots in parallel perspective as luminous dots (single pixels) on a darker background.

A stimulus name consists of the name of the shape followed by the type of rotation (e.g. $u + -0l$), resulting in 108 possible names. Using parallel perspective, there is a fundamental ambiguity with the KDE: reversing the depth values and rotation direction of a particular shape and rotation produces exactly the same display. In other words, a convexity rotating to the right produces exactly the same set of 2D dot motions as a concavity rotating to the left. Thus, $u + -0l$ and $u - +0r$ describe precisely the same display type. There is also no difference in display type among *u*000*l*, *u*000*r*, *d*000*l* and *d*000*r*. This results in a total of 53 distinct display types.

These experiments used 54 white-on-gray dot displays. including two instantiations of the flat stimulus *u*000 (with different dot placements) and one instantiation of each other display type. Each set of dots was windowed to a display area of 182 × 182 pixels (corresponding to the $s \times s$ square), with dots presented as single luminous pixels.

When the dots on the surface of a shape move back and forth in the display, the local dot density changes as the steepness of the hills and valleys changes (with respect to the line of sight). In previous work (Sperling et al., 1989), we showed that this density cue is neither necessary nor sufficient for the perception of depth. However, it is a weak cue which one of three highly trained subjects was able to use for modest above-chance performance when it was presented in isolation. In other words, changing dot density is an artifactual cue to the task. As in previous experiments, we remove this cue by deleting or adding dots as needed throughout the display in order to keep local dot density constant. As a result of this manipulation, all displays had approx. 300 dots visible in the display window. The removal of the density cue results in a small amount of dot scintillation that neither lowers performance substantially nor appears to be useful as an artifactual cue (Sperling et al., 1989, 1990).

Other tokens. The 54 stimuli described so far consisted of luminous dots moving to and fro on a less luminous background. All other stimuli were based upon these displays. First, three conditions involved changes of the token that carried the motion. The moving dots were replaced with disks, patterned disks, or wires. We refer to the dot, wire, and disk conditions as *white-on-gray* stimuli, and the patterned disks as *pattern-on-gray.*

To create a disk stimulus, a dot stimulus is modified in the following way. Each luminous dot in the stimulus is replaced with a 6 × 6 pixel luminous diamond centered on the dot (Fig. 2b), which appears disk-like from the viewing distance used in the experiment. A sample image of white-on-gray disks is depicted in Fig. 2c, and is based on the white-on-gray dot stimulus frame shown in Fig. 2a.

The pattern-on-gray disk stimuli are generated in a similar fashion. The 6 × 6 diamond consists of 24 pixels which are a mixture of black and white (12 of each). These are displayed on an intermediate gray background. The diamond pattern and a sample stimulus frame are shown in Fig. 2d and e, respectively. Note that the diamond pattern has an equal number of black and white pixels in each row.

Other stimuli were based on "wires". Each dot was connected by a straight line (subject to the pixel sampling density) to all neighbors that were at a 2D distance no greater than 15.5 pixels (Fig. 2f). Note that a vector is drawn between two points based on their distance *in the image*, not on their simulated 3D distance. Since the lines were straight, when set in motion they objectively define a thickened surface with lines cutting through the interior of each bump and concavity. This may have yielded a perceived

Fig. 2 (*opposite*). Stimulus display generation for expt 1. (a) A single frame of a white-on-gray dots stimulus. All displays shown in this figure are based on this stimulus frame. (b) The diamond shape used to generate the disks from the dots. (c) A white-on-gray disks stimulus frame. (d) The patterned diamond for the pattern-on-gray condition. (e) A pattern-on-gray frame. (f) A white-on-gray wires frame. All pairs of dots in Fig. 2A were connected whose inter-point distance was less than 15.5 pixels. (g) A frame of dynamic-on-gray dots. In this condition each dot was painted black or white randomly and independently with probability of 0.5 for each color. (h) A frame of dynamic-on-gray disks. The same procedure as in (g) was applied to each pixel lying in each disk. (i) A frame of dynamic-on-gray wires. (j) A frame of dynamic-on-static disks. For both dynamic-on-static conditions (disks and wires), the tokens and the background consisted of random dot noise, and so the tokens cannot be discerned from a single static frame. (k) A frame of the pattern-on-static condition. This frame contains 300 copies of the pattern in (d) on a static noise background. The camouflage is quite effective. (l) An enlargement of the central portion of (k), with the patterned disks emphasized.

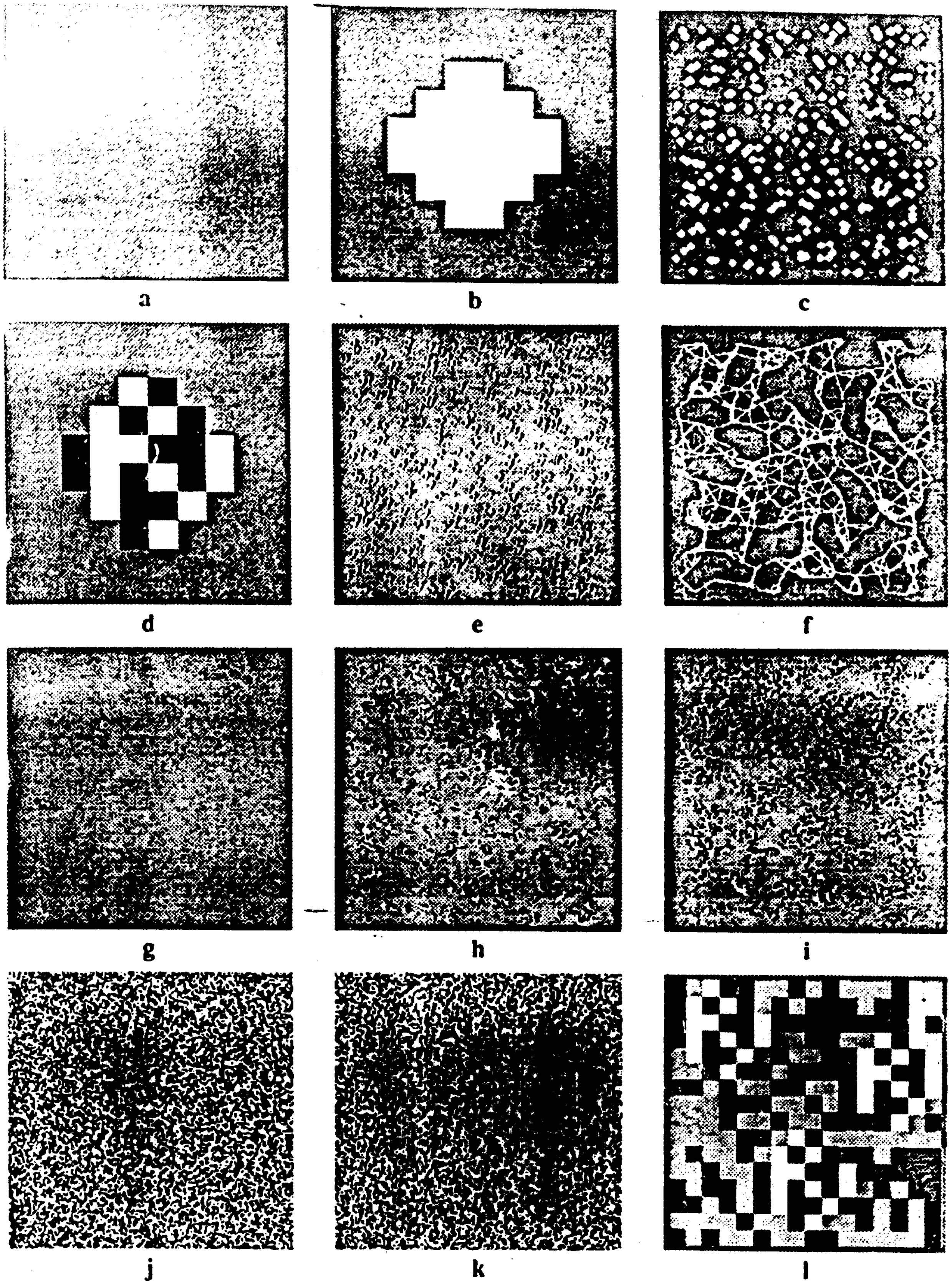


Fig. 2

(tesselated) surface having slightly less relative depth than the base surface. The choice of 15.5 pixels as the criterion for drawing a line was a compromise set in order to make sure that all stimulus dots became an endpoint to at least one line, and that no line was so long as to excessively cut through the simulated surface.

The white-on-gray disks and pattern-on-gray disks were based on the dot stimuli. The same exact instantiations were used in all three conditions. The *n*th frame of a given shape and rotation consisted of either dots, disks or patterned disks centered on the same set of image positions. For the wire stimuli, a new set of 54 instantiations was made.

Dynamic-on-gray. Three types of stimuli were used to explore the motion of patches of dynamic noise moving on a gray background. These stimuli are microbalanced, as we discussed in the previous section. These stimuli are derived from the dot, disk, and wire stimuli. To produce a dynamic-on-gray stimulus from a white-on-gray stimulus, simply change the luminance of each white pixel in each stimulus frame (i.e. the foreground or token pixels) to black randomly and independently with probability 0.5. Thus, foreground pixels undergo random contrast polarity alternation while background pixels are gray (i.e. have zero contrast). Sample frames are illustrated in Fig. 2g, h and i.

Dynamic-on-static. Two types of stimuli were used to explore the motion of patches of dynamic noise moving on a static noise background. This class of stimuli is also microbalanced (Chubb & Sperling, 1988b). We derive dynamic-on-static stimuli from the disk and wire stimuli. The foreground pixels consist of dynamic noise, just as in the previous dynamic-on-gray case. The background pixels consist of a static frame of patterned texture, where each pixel is randomly chosen to be either black or white with a probability of 0.5, just as the dynamic noise is. If a given pixel is a background position for two successive frames, then its color does not change. If that position is a foreground pixel in either or both frames, then there is a 50% chance that its color will change. A single frame of dynamic-on-static stimulus is simply a frame of random dot noise (Fig. 2j). The motion-carrying tokens are not discernible from a single frame. Rather, the areas of moving dynamic noise define the foreground tokens.

Contrast polarity alteration. Three stimulus conditions involved contrast polarity alternation. This stimulus manipulation was explored thoroughly for dot stimuli in the preceding paper (Dosher et al., 1989b). In this condition, the motion-carrying tokens alternate from white to black to white again on successive frames, all against a background of intermediate gray. Constrast polarity alternation was used with dots, disks, and wires, resulting in three polarity alternation conditions.

Pattern-on-static. The final condition involves pattern camouflage. This condition is derived from the pattern-on-gray stimuli. The gray background is replaced with a frame of static random dot noise. In other words, the patterned disk tokens move to and fro in front of a screen of static random dots, occluding it (and occasionally each other) as they pass by. A frame of this stimulus condition is pictured in Fig. 2k, and enlarged in Fig. 2l, where we have artificially highlighted the patterned disks for comparison to the pattern kernel shown in Fig. 2d. There are approx. 300 patterned disks in Fig. 2k. As you can see, the camouflage is quite effective. When the patterned disks move, as one might expect, they are easily visible (Julesz, 1971).

Display details. There are a total of 13 conditions (3 white-on-gray, 1 pattern-on-gray, 3 contrast polarity alternation, 3 dynamic-on-gray, 2 dynamic-on-static, and 1 pattern-on-static). There were 54 distinct displays for each of the 13 conditions. In all conditions, the displays are windowed to an area of 182 × 182 pixels. Displays were computed using the HIPS image processing software (Landy, Cohen & Sperling, 1984a, b), and displayed by an Adage RDS-3000 image display system.

Subjects MSL and JBL viewed these stimuli on a Conrac 7211C19 RGB color monitor. Only the green gun was used, and so stimuli appeared as bright green and black pixels (as dots, disks, lines or noise) on a green background of intermediate luminance. The stimuli subtended 3.7 × 4.2 deg. Stimuli were viewed monocularly through a dark viewing tunnel, using a circular aperture which was slightly larger than the stimuli.

Subject LJJ viewed the stimuli on a US Pixel PX15 black and white monitor with a P4-like phosphor. Here, stimuli subtended 2.9 × 2.9 deg, and appeared as white and black pixels on an intermediate gray background. Stimuli were viewed monocularly through a circular aperture in cardboard which approximately matched the hue of the displays, and which had approximately the same luminance as the stimulus background.

Each stimulus consisted of 30 stimulus frames. These were presented at a 60 Hz frame rate. Each frame was repeated four times, resulting in an effective rate of 15 new stimulus frames per second. Each stimulus lasted 2 sec. A trial sequence consisted of a fixation spot, a blank interval, the 30 frame stimulus, and a blank. The fixation and blank lasted either for 1 sec each (subjects MSL and JBL), or 0.5 sec each (subject LJJ). The background luminance remained constant throughout the trial sequence. Subjects were free to use eye movements to actively explore the display. Stimuli were viewed from a distance of 1.6 m. After each stimulus display, subjects responded with the name of the shape and rotation direction using either a computer keyboard or response buttons.

Slightly different image luminances were used for each subject. The background luminance for subjects MSL, JBL and LJJ were 31.0, 40.0 and 45.0 cd/m^2 respectively. Since isolated luminous pixels were used, the appropriate unit of measurement is *extra* μcd/pixel for bright pixels, and *removed* μcd/pixel for dark pixels, all at a specified viewing distance (Sperling, 1971). Stimuli were calibrated so that extra μcd/pixel and removed μcd/pixel were equal. For subjects MSL, JBL and LJJ, these were 13.2, 19.2 and 15.7 μcd/pixel, respectively, at a viewing distance of 1.6 m. Contrasts were nominally 100%.

Procedure. There were 13 stimulus conditions. For each condition, there were 54 stimuli (two instantiations of the flat stimulus *u*000, and one instantiation of each of the 52 other possible distinct shape/rotation combinations). This resulted in 702 stimuli, each of which was viewed once by each subject. These 702 trials were viewed in random order in six blocks of 117 trials. On a given trial, a stimulus was shown, subjects keyed in their responses, and then feedback was provided so that we measured the best performance of which the subject was capable. Each block lasted approx. 1 hr. Subjects ran several practice sessions on the white-on-gray dots condition before data were collected. Given the mix of stimuli in a given condition, guessing base rates for the identification of shape and rotation direction were between 1/53 (for a strategy of random guessing) and 2/54 (for a strategy of always answering *u*000*l*, or one of its equivalents).

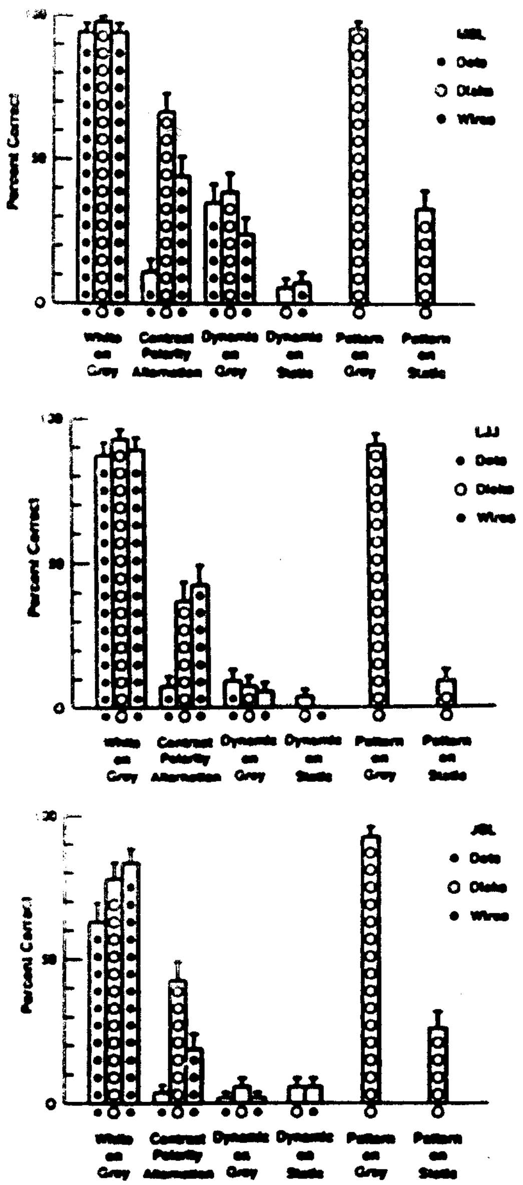


Fig. 3. Results of expt 1. Results are given for three subjects. Different symbols in the bars represent different tokens (large open dots for the disk and patterned disk tokens, small solid dots for the dot tokens, and asterisks for the wire tokens).

Results

The results for the three subjects are summarized in Fig. 3. Each performance measure given here is the percent correct over 54 trials. We discuss each class of stimulus condition in turn.

White-on-gray/Pattern-on-gray. As expected, the performance on the three white-on-gray and the one pattern-on-gray condition was uniformly high. The tokens provided excellent motion signals because they were moving rigid areas of high contrast. It did not particularly matter whether we used dots, as in our previous studies, wires, as in the early wire-frame KDE work (Wallach & O'Connell, 1953), disks, or patterned disks. The disk and patterned disk stimuli provided very strong percepts of shape, although the disks did not undergo realistic foreshortening as they rotated. In fact, the dot stimuli gave the weakest percept of depth. These tokens had the least contrast energy (i.e. were the smallest), and hence were harder to detect. Subject JBL had the greatest difficulty in seeing these small dots, and his results show a slight drop in performance for the dot stimuli.

Dynamic-on-gray. The motion of a token filled with dynamic random dot noise moving on a gray background is microbalanced. In other words, 1st-order motion detectors are "blind" to this stimulus. The expected value of the output of such a detector is zero (across random realizations of the stimulus). Simple 2nd-order mechanisms (e.g. using rectification) serve to reveal the true motion.

The results for three subjects are somewhat different. For two subjects (LJJ and JBL), performance is always at or near chance (less than 10% correct in all cases), although for subject LJJ with the dynamic-on-gray dots the performance is significantly above chance ($P < 0.05$). On the other hand, for subject MSL, performance is always well above chance (24–39% correct identifications), but far less than his nearly perfect (94–98% correct) performance with white or pattern tokens on gray.*

The 1st-order motion mechanisms are clearly the most effective input to the KDE system, since eliminating motion detectable by 1st-order mechanisms reduces performance substantially for all subjects. The results for subject MSL suggest that 2nd-order motion mechanisms can also be used. On some trials, fragments of the microbalanced stimuli did appear 3D to this subject (one of the authors), especially in the foveally-viewed portion of the stimulus. To raise his performance level, he used sophisticated guessing strategies based on active eye movements and local measurements of motion or three-dimensionality in the fovea at a small number of locations of the display. But, these strategies only serve to bring performance up to mediocre levels in comparison with performance with rigid white-on-gray motion.

Dynamic-on-static. The dynamic-on-static manipulation also results in a micro-balanced stimulus. For the dynamic-on-static conditions, performance is at chance level for all three subjects, and for both wire disk tokens. As with the dynamic-on-gray conditions, the motion of the tokens is visible. It is not particularly difficult to detect the motion of an area of dynamic noise on a static noise background (Chubb & Sperling, 1988b). However, this sort of motion engenders no shape percept whatever under the conditions of our experiments.

Unlike dynamic-on-gray stimuli, dynamic-on-static stimuli are not revealed by contrast rectification. Detection of the motion of a region of flicker requires more elaborate 2nd-order mechanisms. Regions of flicker could first be detected by applying a linear temporal filter (such as differentiation), followed by rectification, and then by application of a 1st-order motion mechanism. Some such complex 2nd-order motion detector exists in the human visual system, since we are capable of seeing areas of flicker move, including in the displays of our experiment (at least with scrutiny). Yet, this 2nd-order motion detection system does not support the structure-from-motion computation for our dynamic-on-static stimuli.

Prazdny (1986) reached the opposite conclusion using dynamic-on-static displays representing simple wire objects rotating in a tumbling motion. Each object contained five wires, and subjects were required to identify the object among six alternative wire-frame objects.

*In order to test the range of luminances over which polarity alteration was effective, we ran a control experiment (using MSL and JBL as subjects), where a variety of white pixel luminances were used with a given black pixel luminance. We viewed a variety of dynamic-on-gray displays, varying the luminance values for the black and white pixels independently over a wide range. We also tested a variety of other luminance calibration procedures. Dynamic-on-gray stimuli are only micro-balanced if the contrast energy of the white pixels is the same as that of the black pixels. And, it is difficult to calibrate the luminance of individual pixels embedded in a complex display texture given that the desired pattern is first low-pass filtered by the CRT video amplifier, and then passes through the gun nonlinearity (see Mulligan & Stone, 1989, for a full discussion of this point). Thus, it was important to verify that our results were robust over a range of luminance values overlapping the calibrated equal contrast point.

To summarize, shape identification performance is consistent with the results of expt 1 for a reasonably wide range of white pixel luminances. Subject MSL consistently performs at moderate levels, and subject JBL consistently performs at or near chance. The luminance levels yielding poor shape identification performance are consistent with the levels that result in the weakest 3D percept, and are roughly consistent with the luminance levels that are balanced (black pixel decrement vs white pixel increment) for a variety of calibration displays. The performance levels for dynamic-on-gray stimuli in expt 1 do not result from a miscalibration of luminance levels.

The displays were 7 × 7 deg, and the wires were several pixels thick. Performance was quite high in the task for five subjects. Although we have some reservations about the experimental method employed by Prazdny, we have generated similar displays in our laboratory, and our dynamic-on-static wire-frame displays do yield a shape percept when displays are restricted to a small number of wires.

The most likely explanation of the difference between our results and those of Prazdny involves the difference in spatial resolution required by each task. Chubb and Sperling (1988a) have demonstrated that 2nd-order motion systems have less spatial resolution than the 1st-order mechanisms, and that their resolution drops precipitously with increases in retinal eccentricity. In our displays, motion was about a vertical axis using parallel perspective, and hence all motion was along the horizontal. There could be as many as 10 or 20 disks or wires in a given row of the image to resolve. Our displays did not yield a global percept of optic flow, but motion was perceived foveally with scrutiny. This is entirely consistent with Chubb and Sperling's observation. Prazdny did not give precise details about his stimuli, but it was clear that along a given motion path there were only two or three wires to resolve across his far larger display. Performance was so low in our dynamic-on-static conditions because too much spatial acuity was required of the 2nd-order system that detects the motion of flickering regions.

How useful for perception of shape is a display of dynamic noise figures moving on a static noise background? We have examined a large number of disk and (thick) wire displays in order to span the gap of spatial resolution between Prazdny's displays and our own. With our 3 × 3 deg display size, a shape percept can only be achieved by using a very small number of tokens (around 5–10). These displays consisted of rotating disk tokens. Cavanagh and Ramachandran (1988) suggest an alternative explanation of the difference between our results and those of Prazdny. They consider the crucial difference to be that the objects portrayed in the Prazdny displays were connected (one long wire figure), whereas our displays consisted of separate disk tokens. With our wire displays, almost no 3D percept was achieved for the dynamic-on-static condition. In addition, we were able to achieve a 3D percept with displays of a small number of dynamic-on-static disks. Thus, we feel that low spatial resolution in the 2nd-order motion system (rather than unconnected tokens) is the likely explanation for failure of KDE.

Contrast polarity alternation. Performance is quite poor for the contrast polarity-alternating dots as it was in the previous paper (Dosher et al., 1989b). For two subjects (JBL and LJJ) performance is at chance or insignificantly above chance. For subject MSL, performance is low (11% correct) but significantly above chance ($P < 0.05$). On the other hand, when the token is changed to disks or wires, performance rises substantially. Contrast polarity alternation is not as devastating a stimulus manipulation for disks and wires as it is for dots.

For 1st-order motion detection mechanisms such as the Reichardt detector, contrast polarity alternation causes the strongest responses to be in the wrong direction. Yet, the intended motion can be detected quite accurately if a 2nd-order detector is used that first applies a luminance nonlinearity followed by a Reichardt detector. The primary difference between the dots on the one hand, and the disks and wires on the other, is that the disks and wires have more pixels illuminated. In other words, they have more contrast energy, and in particular thay have more energy at lower spatial frequencies. Thus, the disk and wire stimuli should stimulate both the 1st- and 2nd-order motion detection systems more strongly, resulting in stronger incorrect direction information from the 1st-order system as a whole, but also stronger information from the 2nd-order system, and stronger directional information in those selected 1st-order frequency bands which signal the correct direction.

It is interesting to note that a large number of the errors made by observers with polarity-alternating stimuli were errors in the direction of rotation *only*, with the shape specified correctly. For example, for a stimulus which had as correct answers either $u+-0l$ or $u-+0r$, the subject incorrectly responded with $u+-0r$ or $u-+0l$, rather than with any of the 104 other possible incorrect responses. This effect was largest for the disk tokens. In a separate control experiment, for contrast polarity-alternating disk stimuli, 39% of the errors made by subject MSL were only an error in the specification of direction, compared to 1.4% direction errors for the dynamic-on-gray conditions. For subject JBL, the corresponding values were 48% and 5.6%. For the polarity-alternating disks, on trials when subject MSL correctly identified the shape, there was a 33% chance that he would misidentify the direction of rotation (for JBL: 29.3%). We believe that accurate shape identification in this condition primarily reflects responses constructed from selected 1st-order information. One strategy was simply to specify the opposite rotation direction to that which was perceived! The displays did, however, occasionally appear to be 3D with the correct direction of motion (at certain times during the rotation, or close to the location to which the eyes were directed), indicating a residual 2nd-order motion input to the KDE system. The fact that these displays only appeared foveally to be rotating in the correct direction, and then only using the larger tokens, is consistent with a 2nd-order motion detection system with low contrast sensitivity and low spatial resolution (as has been demonstrated by Chubb & Sperling, 1988b), and more sensitive in the fovea (Chubb & Sperling, 1988a). In summary, we have some indication that 2nd-order motion detection mechanisms can be used to derive 3D structure, but they are far less robust and have poorer spatial resolution than 1st-order motion mechanisms.

Pattern-on-static. For all three subjects performance with pattern-on-static displays is quite poor (9, 26 and 33% correct), although it is significantly above chance levels in all cases ($P < 0.05$). This poor performance results from a mismatch of resolution and temporal sampling. The patterned disks are quite detailed/high frequency. The disks are 6 pixels in diameter, and can move as far as 8.3 pixels in one frame. This speed is only achieved by disks at the top of a peak when in the middle of the display (i.e. near frame numbers 0, 15 and 29), but many disks are moving 3–5 pixels per frame. High frequency spatial filters which are required to identify the disks must correlate across frames with filters that are far more than 90 deg away in the phase of their peak spatial frequency. A typical 1st-order detector will not compare spatial regions that far apart in order to avoid spatio-temporal aliasing (van Santen & Sperling, 1984). Thus, the clearest motion signals are coming from the slower areas in the display, which are the least useful for discriminating the shapes. We have examined pattern-on-static displays with finer temporal sampling (60 new frames per sec, as opposed to 4 repaints of 15 new frames per sec used in the experiment), and they give a strong impression of three-dimensionality. Thus, poor performance in the task resulted from undersampling in time of the stimuli, which interferes with 1st-order (and some 2nd-order) motion mechanisms, and good KDE can result from the motion of tokens which are camouflaged when at rest.

We have also examined dynamic-on-static displays with finer temporal sampling (60 new frames per sec). These displays yield no impression of three-dimensionality. The poor results for dynamic-on-static displays do not result from insufficient sampling in time. Also, since finely sampled pattern-on-static displays do appear 3D, poor performance with dynamic-on-static-displays does not result from the camouflage of the tokens when at rest. Rather, dynamic-on-static displays yield no effective KDE because of the low resolution of the 2nd-order system required to analyze the motion.

EXPERIMENT 2. TWO-FRAME KDE

The first experiment shows that accurate performance in shape identification is dependent upon a global (primarily 1st-order) optic flow. If a stimulus manipulation makes that optic flow noisy or otherwise interferes with the optic flow computation, there is little or no KDE. This occurs even though foveal scrutiny does reveal the motion in these displays.

If the percept of surface shape depends upon a global optic flow, then we should be able to get reasonable shape identification performance from any stimulus that results in a strong percept of optic flow. In particular, the extended (2 sec) viewing conditions of expt 1 should not be necessary. Two frames are obviously the minimum number of frames that can yield a percept of motion, and two frames should suffice. In the second experiment, we investigate the accuracy of performance in the shape identification task for two-frame displays.

Method

Subjects. There were two subjects in this experiment. One was an author, and the other was a graduate student naive to the purposes of this experiment. Both had normal or corrected-to-normal vision. There were slight differences in the conditions for each of the two subjects. These will be pointed out below.

Stimuli and apparatus. The stimuli were similar to the white-on-gray dot stimuli from expt 1. Stimuli were generated from the same set of 3D shapes, using the same dot densities, and projected in the same way. The local dot density was kept constant using the same scintillation procedure. New stimuli were computed, two of the flat shape, and one of each of the other 52 shapes, resulting in 54 displays.

Each display consisted of 11 frames, rotating from 20 deg left to 20 deg right in increments of 4 deg per frame. The middle frame (number 6) was face-forward, as was the first frame of each display in expt 1. Two-frame stimuli consisted of a presentation of the middle frame followed by one of the other 10 display frames. This resulted in either a leftward or rightward rotation of 4–20 deg between the two frames of the display. A single trial display consisted of 0.5 sec of a cue spot, 0.5 sec blank, the first frame, an inter-stimulus blank interval (or ISI), the second frame, and a blank. Each stimulus frame was repainted four times at 60 Hz, for a total duration of 67 msec. We define the ISI to be the time interval between the onset of the last painting of the first stimulus frame and the onset of the first painting of the second stimulus frame. For example, when no blank frames were used, the ISI was 16.7 msec. Displays were 182 × 182 pixels, and were presented using the same apparatus and viewing conditions as for subject LJJ in expt 1. The background luminances for subjects MSL and LJJ were 15.6 cd/m^2 and 5.0 cd/m^2, respectively. The corresponding dot luminosities were 26.8 and 15.7 extra μcd/dot, respectively. Nominal contrasts were huge (i.e. nominal Weber contrasts of 500% or more).

Procedure. The task was shape and rotation identification. Subjects keyed their responses using response buttons, and received feedback on the display after their response. Three groups of trials were run. In the first, the ISI was 16.7 msec, and rotation angle between frames was varied from 4 to 20 deg. Since the second frame could be chosen from either the frames preceding or succeeding the middle frame (rotation to the left or right), this resulted in 540 possible stimuli (54 displays, 2 directions, 5 rotation angles). These were run in random order in 4 blocks of 135 trials. In the second group of trials, rotation was kept constant at 4 deg. ISI ranged from 16.7 to 83.3 msec. This again resulted in 540 trials presented in random order in 4 blocks of 135 trials. In the third group

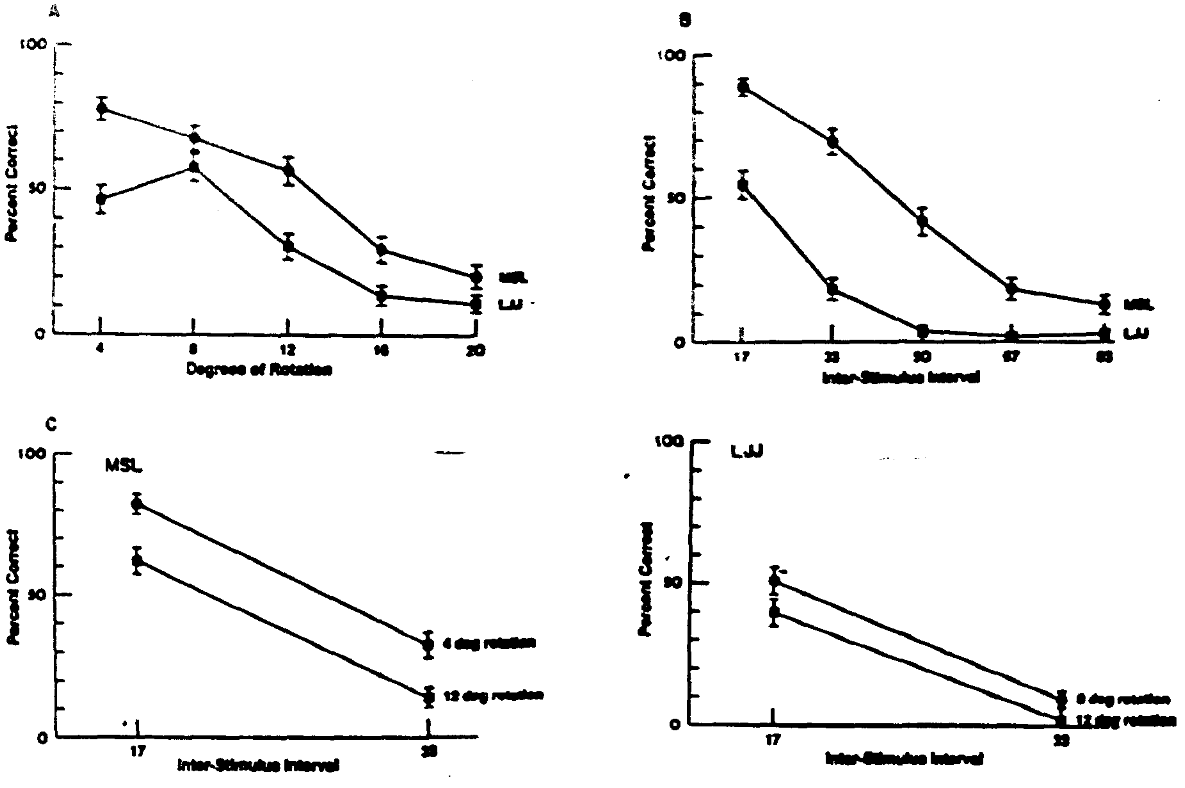


Fig. 4. Results of expt 2. Data for two subjects are shown. Error bars indicate ±1 SEM. (A) Shape-and-rotation identification accuracy as a function of the angle of rotation between the two frames. ISI was 16.7 msec. (B) Shape-and-rotation identification accuracy as a function of the duration of a blank inter-stimulus interval (ISI). Rotation angle was 4 deg. (C) The two manipulations used in the same experiment. Note the lack of interaction.

of trials, both rotation angle and ISI were varied. The ISIs were either 16.7 or 33.3 msec. For subject MSL, the rotation angles were either 4 or 12 deg. For LJJ, they were either 8 or 12 deg. These four conditions (two rotation angles by two ISIs) resulted in 432 trials which were presented in random order in 4 blocks of 108 trials.

Results

The results are shown in Fig. 4. Each data point is the percent correct over 108 trials. As is evident from the figure, shape identification can be quite high for these minimal motion displays (for similar observations using different experimental methodology, see Braunstein, Hoffman, Shapiro, Andersen & Bennett, 1987; Lappin, Doner & Kottas, 1980; Mather, 1989; and Petersik, 1980). For an ISI of 16.7 msec (Fig 4A), this entire sequence lasted only 133 msec. Yet, performance was as high as 54.6% for subject LJJ, and 88.9% for subject MSL (62.8% and 94.2% of their white-on-gray dots performance in expt 1, respectively). Two frames of moving dots are sufficient for accurate, although not perfect performance in this shape identification task. Since these experiments were first reported (Landy, Sperling, Dosher & Perkins, 1987a; Landy, Sperling, Perkins & Dosher, 1987b), Todd (1988) has also shown above-chance KDE performance for two-frame stimuli, although in his paradigm the two frames are repeated several times before a response is made.

Rotation angle and fixation. Performance as a function of rotation angle between the two frames is given in Fig. 4A. Performance decreases with increasing angle of rotation for subject MSL. For subject LJJ, performance reaches a peak at 8 deg, and decreases for smaller and larger rotations. The decrease in performance with larger rotation angles is to be expected, since the correspondence problem becomes increasingly difficult as dots move farther from their initial positions. One might also expect performance to drop as rotation angle decreases to zero. At extremely small rotation angles, the remaining motion would fall below threshold. In our displays, the drop with small rotation angles might be expected to occur even sooner as the small motions in the display became corrupted by poor spatial sampling (inter-pixel distance was approx. 1 min arc). This drop was only seen in the data of LJJ, and presumably would be seen in those of MSL if he had been tested using smaller rotations.

In a previous paper (Dosher et al., 1989b), we found that adding a blank interval between successive frames of a 30 frame KDE stimulus reduced shape identification to near chance performance. This was explained by reduction of power in the stimulus to the 1st-order system. This effect is also seen here, where performance decreases monotonically with increasing ISI (Fig. 4B). Subject LJJ performs at chance levels with a 50 msec or greater ISI, while subject MSL is still slightly above chance performance with an 83.3 msec ISI.

Time and distance. In the previous two groups of trials, there was a confounding between the stimulus manipulation (rotation angle or ISI) and dot velocity. Greater rotation angles at a fixed (16.7 msec) ISI produced greater velocities. Similarly, greater ISIs at a fixed 4 deg rotation angle resulted in smaller velocities. If performance were simply a function of velocity, then rotation angle and ISI should trade off. In Fig. 4C we present the results of varying both ISI and rotation angle factorially. We used a different set of rotations for subject LJJ than MSL based on the results in Fig. 4A, so that for both subjects the performance was expected to decrease with increasing rotation angles. As can be seen in the figure, the two variables do not trade off as would be expected if performance were only a function of velocity, or rotation speed. Increasing rotation angle increases the difficulty of the correspondence problem. Increasing ISI causes increasing problems for the motion detection system. Both manipulations degrade performance in an additive fashion. This observation contradicts Korte's (1915) 3rd law of apparent motion perception, which states that an increase in ISI must be counteracted by an increase in distance traveled for strong apparent motion. In Fig. 4C, Korte's law predicts a cross-over interaction, which is strongly disconfirmed. However, Burt and Sperling (1981) show that time and distance have independent additive effects on the strength of the apparent motion of dot stimuli, which agrees with the present results.

KDE from optic flow. Accurate KDE performance requires a global optic flow. When that optic flow is produced by a minimal motion stimulus—a two-frame display—the shape percept may be fragile and easily degraded by a variety of stimulus manipulations. The stimuli are quite brief in this paradigm and, by subject reports, appear as a collection of dots moving at various speeds, i.e. "look like" an optic flow. On some trials, only patches of planar motion are perceived, and the shape response is generated cognitively. On other trials, a 3D surface is perceived. On some trials the optic flow is perceived and so is the shape, but the shape percept is only "felt" after the display is over. As we discussed extensively in our first article on the shape identification task (Sperling et al., 1989), KDE is inextricably tied with the percept of an optic flow. It can be very difficult to differentiate empirically between a judgment based on a 3D percept and performance based on an alternative strategy (computationally equivalent to that required for KDE) using a remembered set of 2D velocities.

Reasonably accurate performance on the shape-and-rotation identification task results from only two frames of 300 points. In the computer vision literature, there have been several studies of the structure-from-motion problem resulting in theorems of the following form: "*m* views of *n* points under the following restrictions of the motion path suffice to determine the 3D structure up to a reflection" (Bennett & Hoffman, 1985; Hoffman & Bennett, 1985; Hoffman & Flinchbaugh, 1982; Ullman, 1979). It has been suggested that these minimal conditions for structure from motion also govern human perception (Braunstein et al., 1987; Petersik, 1987). The particular models just mentioned do not have any prediction concerning performance in the 300 points/2 views situation used here. An exception is a recent paper by Bennett, Hoffman, Nicola and Prakash (1989), where it is shown that there is a one parameter family of possible interpretations for two frames of four or more points. This family is parameterized by the slant of the axis of rotation (as in the "isokinescopic displays" described by Adelson, 1985), and the paper does not deal explicitly with rotation axes in the image plane, as used here. On the other hand, models that compute 3D structure based only upon a single velocity field do allow for this performance (Longuet-Higgins & Prazdny, 1980; Koenderink & van Doorn, 1986). We take our experimental results as evidence for optic flow-based methods for the KDE, as opposed to models requiring three or more views. In particular, our results strongly rule out models that require measurement of acceleration in addition to velocity (e.g. Hoffman, 1982).

Structure-from-motion computation may improve its 3D representation with additional information (e.g. with additional frames, Grzywacz, Hildreth, Inada & Adelson, 1988; Hildreth & Grzywacz, 1986; Landy, 1987; Ullman, 1984). The shape in our two-frame displays does not [illegible] appear to have the depth extent that [illegible] from the 30 frame displays of expt 1, and two-frame performance is reduced relative to 30-frame performance. The shape identification task can be solved by knowing only the sign of depth and direction of motion in each spatial location (up to a reflection), without accurately estimating either velocity or the amount of depth.

DISCUSSION

Two experiments investigated the type of motion detection mechanism used as an input to the structure-from-motion system. Performance in the shape-and-rotation identification task was accurate regardless of the token used to carry the motion, as long as that token was presented with constant contrast polarity (the white-on-gray and pattern-on-gray conditions). The performance decrements seen with contrast polarity alternation and the two microbalanced conditions add further evidence to the conclusion of Dosher et al. (1989b) that 1st-order motion detectors are the primary substrate for the computation of shape. In addition, there are indications of an input to the shape computation from 2nd-order motion mechanisms, which is weak, low in spatial resolution, and concentrated at the fovea. 2nd-order mechanisms that require temporal filtering (i.e. detection of flicker) prior to a point nonlinearity were useless here because of the spatial resolution required by our stimuli. These sorts of detectors would only be useful for KDE displays involving a small number of moving features, rather than the densely sampled optic flows required for the determination of precise shapes of curved surfaces from motion cues. The results from the two-frame experiments reinforced these conclusions. They also demonstrated that detection of instantaneous velocity is sufficient for KDE; acceleration is not required, nor are more than two views.

Acknowledgements—The work described in this paper was supported primarily by a grant from the Office of Naval Research, grant N00014-85-K-0077, and partly by USAF Life Science Directorate, grants 85-0364, 88-0140, and NSF grant IST-8418867. We would like to thank Charles Chubb for his helpful comments, and Robert Picardi for technical assistance. Portions of this work have been presented at the annual meetings of the Association for Research on Vision and Ophthalmology, Sarasota, Florida (Landy et al., 1987a) and the Optical Society of America, Rochester, New York (Landy et al., 1987b).

REFERENCES

Adelson, E. H. (1985). Rigid objects appear highly nonrigid. *Investigative Ophthalmology and Visual Science* (Suppl.), *26*, 56.

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*, 284–299.

Anstis, S. M. & Rogers, B. J. (1975). Illusory reversal of depth and movement during changes of contrast. *Vision Research*, *15*, 957–961.

Bennett, B. M. & Hoffman, D. D. (1985). The computation of structure from fixed-axis motion: Nonrigid structures. *Biological Cybernetics*, *51*, 293–300.

Bennett, B. M., Hoffman, D. D., Nicola, J. E. & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America A*, *6*, 1052–1069.

Braunstein, M. L., Hoffman D. D., Shapiro, L. R., Andersen, G. J. & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 335–343.

Burr, D. C., Ross, J. & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London, B*, *227*, 249–265.

Burt, P. & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, *88*, 171–195.

Cavanagh, P. & Ramachandran, V. S. (1988). Structure from motion with equiluminous stimuli. Paper presented to the *Annual Meeting of the Canadian Psychological Association*, Montreal, June.

Chubb, C. & Sperling, G. (1988a). Processing stages in non-Fourier motion perception. *Investigative Ophthalmology and Visual Science* (*Suppl.*), *29*, 266.

Chubb, C. & Sperling, G. (1988b). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, *5*, 1986–2007.

Chubb, C. & Sperling, G. (1989a). Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proceedings of the National Academy of Sciences, U.S.A.*, *86*, 2985–2989.

Chubb, C. & Sperling, G. (1989b). Second-order motion perception: Space/time separable mechanisms. *Proceedings: Workshop on visual motion* (pp. 126–138). Washington, D.C.: IEEE Computer Society Press.

Dosher, B. A., Landy, M. S. & Sperling, G. (1989a). Ratings of kinetic depth in multi-dot displays. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 816–825.

Dosher, B. A., Landy, M. S. & Sperling, G. (1989b). The kinetic depth effect and optic flow—I. 3D shape from Fourier motion. *Vision Research*, *29*, 1789–1813.

Grzywacz, N. M., Hildreth, E. C., Inada, V. K. & Adelson, E. H. (1988). The temporal integration of 3-D structure from motion: A computational and psychophysical study. In von Seelen, W., Shaw, G. & Leinhos, U. M. (Eds.), *Organization of neural networks*. New York: VCH.

Heeger, G. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A*, *4*, 1455–1471.

Hildreth, E. C. & Grzywacz, N. M. (1986). The incremental recovery of structure from motion: Position vs velocity based formulations. *Proceedings of the workshop on motion: Representation and analysis*. IEEE Computer Society no. 696, Charleston, South Carolina, 7–9 May.

Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America 72*, 888–892.

Hoffman, D. D. & Bennett, B. M. (1985). Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America A*, *2*, 350–353.

Hoffman D. D. & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics*, *42*, 195–204.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: The University of Chicago Press.

Koenderink, J. J. & van Doorn, A. J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America A*, *3*, 242–249.

Korte, A. (1915). Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie*, *72*, 193–206.

Landy, M. S. (1987). A parallel model of the kinetic depth effect using local computations. *Journal of the Optical Society of America A*, *4*, 864–876.

Landy, M. S., Cohen, Y. & Sperling, G. (1984a). HIPS: A Unix-based image processing system. *Computer Vision, Graphics and Image Processing*, *25*, 331–347.

Landy, M. S., Cohen, Y. & Sperling, G. (1984b). HIPS: Image processing under UNIX—Software and applications. *Behavior Research Methods, Instruments and Computers*, *16*, 199–216.

Landy, M. S., Sperling, G., Dosher, B. A. & Perkins, M. E. (1987a). Structure from what kinds of motion? *Investigative Ophthalmology and Visual Science* (*Suppl.*), *28*, 233.

Landy, M. S., Sperling, G., Perkins, M. E. & Dosher, B. A. (1987b). Perception of complex shape from optic flow. *Journal of the Optical Society of America A*, *4*, 108.

Lappin, J. S., Doner, J. F. & Kottas, B. L. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science*, *209*, 717–719.

Lelkens, A. M. M. & Koenderink, J. J. (1984). Illusory motion in visual display. *Vision Research*, *24*, 1083–1090.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, *208*, 385–397.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, *211*, 151–180.

Mather, G. (1989). Early motion processes and the kinetic depth effect. *The Quarterly Journal of Experimental Psychology*, *41A*, 183–198.

Mulligan, J. B. & Stone, L. S. (1989). Halftoning method for the generation of motion stimuli. *Journal of the Optical Society of America A*, *6*, 1217–1227.

Petersik, J. T. (1980). The effects of spatial and temporal factors on the perception of stoboscopic rotation simulations. *Perception*, *9*, 271–283.

Petersik, J. T. (1987). Recovery of structure from motion: Implications for a performance theory based on the structure-from-motion theorem. *Perception and Psychophysics, 42,* 355–364.

Prazdny, K. (1986). Three-dimensional structure from long-range apparent motion. *Perception, 15,* 619–625.

Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research, 13,* 1399–1401.

Reichardt, W. (1957). Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. *Zeitschrift Naturforschung B, 12,* 447–457.

van Santen, J. P. H. & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, 1,* 451–473.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2,* 300–321.

Sperling, G. (1971). The description and luminous calibration of cathode ray oscilloscope visual displays. *Behavior Research Methods and Instruments, 3,* 148–151.

Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods and Instrumentation, 8,* 144–151.

Sperling, G., Landy, M. S., Dosher, B. A. & Perkins, M. E. (1989). The kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 826–840.

Sperling, G., Dosher, B. A. & Landy, M. S. (1990). How to study the kinetic depth effect experimentally. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 445–450.

Todd, J. T. (1988). Perceived 3D structure from 2-frame apparent motion. *Investigative Ophthalmology and Visual Science (Suppl.), 29,* 265.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and non-rigid motion. *Perception, 13,* 255–274.

Wallach, H. & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45,* 205–217.

Watson, A. B. & Ahumada, A. J. Jr (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, 1,* 322–342.